ChIP-seq Reveals Mutation-Specific Pathomechanisms of *HOXD13* Missense Mutations

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

im Fach Biologie
eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Diplom-Biologe Daniel Murad Ibrahim

Präsidentin/Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Stefan Hecht, PhD

Gutachter/innen:
1. Prof. Dr. med Stefan Mundlos
2. Prof. Dr. rer. nat. Petra Seemann
3. Prof. Dr. rer. nat. Uwe Ohler

Tag der mündlichen Prüfung: Mittwoch, der 16. April 2014

„*I am the eggman, they are the eggman,*

I am the walrus, goo goo goo joob."

Lennon/McCartney (1967)

Table of Contents

Summary

The correct regulation of gene expression in space and time is central to embryonic development and is mediated by transcription factors (TF). Transcription factors are DNA-binding proteins that control the expression of distinct sets of target genes. Mutations in TFs frequently underlie hereditary diseases, since the mutation not only affects the TF itself, but can also change the expression of the TF's target genes.

The genes of the *HOXA* and *HOXD* clusters are a group of evolutionary highly conserved TFs that control the formation of the anterior-posterior body axis as well as the patterning of the vertebrate limb. More than 20 distinct pathogenic mutations in *HOXD13*, the most posteriorly expressed gene of the *HOXD*-cluster, have been associated with a broad range of disease phenotypes, including synpolydactyly, clinodactyly, and different types of brachydactylies. However, a molecular basis providing a genotype-phenotype correlation for the diverse *HOXD13*-associated diseases remains elusive.

To date, the experimental methods used to functionally characterize *HOXD13* mutations have allowed only limited insights into the underlying molecular pathomechanisms. While ChIP-seq technology has proven to be a powerful new method to investigate control mechanisms of TFs on a genome-wide scale, a number of technical hurdles hinder the application of ChIP-seq for functional characterization of TFs mutations involved in hereditary diseases.

The aim of this study was to overcome these limitations by establishing a new cell-culture based ChIP-seq approach that can be used to investigate a wide spectrum of TFs and TF mutations. The new method was applied to characterize two previously unknown missense mutations in *HOXD13*, p.Q317K and p.R298Q, which both alter the DNA-binding domain of HOXD13 but lead to very different disease phenotypes.

Initial analysis of the HOXD13^{Q317K} mutant indicated an alteration of the TFs recognition sequence, which resembled the recognition sequence of another TF, PITX1. Detailed analyses of ChIP-seq data revealed a genome-wide shift in binding of HOXD13^{Q317K} towards a more PITX1-like binding pattern. Even further, expression analysis and viral overexpression in developing chicken limb buds confirmed that the mutation introduced a partial conversion of HOXD13^{Q317K} into a TF with PITX1-like properties. Examination of the HOXD13^{R298Q} mutant suggested a quantitative instead of qualitative modification of binding activity. Here, initial analysis of genomic binding was followed by detailed comparison between wildtype and mutant binding sites. In combination with the patient phenotype, these analyses point to a dominant-negative effect of the HOXD13^{R298Q} mutation, which might be caused by a more complex biochemical mechanism.

In addition to functional characterization of the two HOXD13 missense mutations, analysis of genome-wide HOXD13 binding revealed a set of binding sites that suggest cooperative gene regulation by HOXD13 and other posterior HOX TFs.

Collectively, applying ChIP-seq to functionally characterize the pathophysiology of HOXD13 mutations provides direct evidence for distinct molecular pathomechanisms underlying the HOXD13^{Q317K} and HOXD13^{R298Q} missense mutations and point to different aspects necessary for correct HOXD13 function.

Zusammenfassung

Der Embryonalentwicklung liegt eine präzise räumliche und zeitliche Regulation der Genexpression zugrunde. Diese Regulation wird durch Transkriptionsfaktoren ermöglicht – DNA-bindende Proteine, von denen jedes ein bestimmtes Repertoire an sogenannten Zielgenen reguliert. Mutationen von Transkriptionsfaktoren führen häufig zu angeborenen Erbkrankheiten, da eine solche Mutation nicht nur den Transkriptionsfaktor betrifft, sondern auch die Expression seiner Zielgene.

Die Gene der *HOXA*- und *HOXD*-Cluster gehören zu einer Gruppe von evolutionär hochkonservierten Transkriptionsfaktoren, welche in Wirbeltieren sowohl die Genexpression entlang der Längsachse des Körpers, als auch entlang der Extremitäten steuern. Über 20 verschiedene Mutation in *HOXD13*, dem Gen des *HOXD*-Clusters welches das am stärksten posteriore Expressionsmuster aufweist, sind bisher in Zusammenhang mit humanen Erbkrankheiten entdeckt worden. Die *HOXD13*-Mutationen führen zu einer Reihe unterschiedlicher Dysplasien der Extremitäten, darunter Synpolydactylie, Clinodactylie und verschiedene Brachydactylien. Eine molekularbiologische Grundlage für die Vielgestaltigkeit der *HOXD13*-Mutationen ist bisher jedoch noch nicht beschrieben worden.

Die experimentellen Methoden welche bisher zur funktionellen Charakterisierung von *HOXD13*-Mutationen verwendet wurden ermöglichten lediglich eine eingeschränkte Interpretation der zugrunde liegenden Pathomechanismen. Die im letzten Jahrzehnt entwickelte ChIP-seq Technologie zeigt sich als leistungsfähige neue Methode, welche es ermöglicht die Funktion eines Transkriptionsfaktors genomweit zu untersuchen. Um diese Technik jedoch zur funktionellen Charakterisierung von mutanten Transkriptionsfaktoren anzuwenden, gilt es einige technische Herausforderungen zu bewältigen.

In dieser Arbeit wurde eine neue, Zellkultur-basierter Methode etabliert, welche genutzt werden kann um eine Vielzahl von Transkriptionsfaktoren und Transkriptionsfaktormutationen zu untersuchen. Die neue Methode wurde anschließend angewandt um zwei zuvor unbeschriebene Punktmutationen in dem Transkriptionsfaktor HOXD13, p.Q317K und p.R298Q, zu charakterisieren. Obwohl beide Mutationen die DNA-bindende Domäne von HOXD13 verändern, verursachen sie in Patienten stark unterschiedliche Phänotypen.

Eine erste Analyse der ChIP-seq Daten für die HOXD13^{Q317K} Mutante zeigte eine Veränderung der Bindungssequenz, welche nun jener eines anderen Transkriptionsfaktors, PITX1, ähnelte. Weiterführende Analysen der genomweiten Bindung von HOXD13^{Q317K} zeigten, dass die Mutation genomweit zu einem veränderten Bindungsprofil führte, welches eher dem Bindungsmuster von PITX1 entspricht. Darüber hinaus zeigten induzierte Genexpression sowie phänotypische Effekte nach viraler Überexpression in Hühnerextremitätenknospen, dass die HOXD13^{Q317K}-Mutante PITX1-ähnliche Eigenschaften aufweist. Im Gegensatz zu dieser qualitativen Änderung der DNA-Bindung deutete die Analyse der HOXD13^{R298Q}-Mutante

vielmehr auf eine quantitative Veränderung der DNA-Bindung hin. Der initialen Analyse der $HOXD13^{R298Q}$-Bindestellen folgte ein detaillierter Vergleich zwischen $HOXD13^{wt}$- und $HOXD13^{R298Q}$-gebundenen Sequenzen. In Kombination mit dem humanen Phänotyp deutet bei dieser Mutante vieles auf einen dominant-negativen Pathomechanismus hin. Die Analyse zeigte des Weiteren, dass möglicherweise komplexe molekularbiologische Mechanismen diesem Effekt zugrunde liegen.

Zusätzlich zur funktionellen Analyse der mutanten HOXD13 Proteine ergab das genomweite Bindeprofil von HOXD13 erste deutliche Hinweise auf ein kooperatives Binden von HOXD13 mit anderen posterioren HOX-Transkriptionsfaktoren.

Zusammengenommen ermöglichte die Verwendung von ChIP-seq für die funktionelle Charakterisierung von HOXD13 Mutationen zum ersten Mal die Erhebung von experimentellen Daten welche beweisen, dass klar unterscheidbare molekularbiologische Mechanismen den $HOXD13^{Q317K}$- und $HOXD13^{R298Q}$-Mutationen zugrunde liegen und diese wiederum auf unterschiedliche Aspekte der normalen HOXD13 Funktion hinweisen.

1 Introduction

In embryonic development, the body with its organ systems, organs, and appendages arises from a single cell in a stereotypic and highly coordinated way. In this process, cells differentiate into hundreds of different cell types, all carrying the same DNA. Consequently, cells do not differ in their genetic information but rather in the way they use it. Therefore, every cell is determined by the genes it expresses.

The regulation of gene expression is mediated by transcription factors (TF). Each TF regulates the expression of a distinct set of target genes and each cell expresses a unique combination of TFs that is referred to as the cell's regulatory state (Davidson 2006). Mutations in TFs not on ly affect the TF itself, but also the expression of its target genes. Misexpression of genes due to TF mutations severely affects the developing organism, since correct embryonic development results from a series of highly coordinated gene expression steps. Therefore, it is of little surprise that TF mutations are frequently underlying hereditary diseases of congenital malformations (Boyadjiev and Jabs 2000; Vaquerizas et al. 2009).

1.1 Transcription Factors

In the vertebrate genome, 6-10% of all protein-coding genes code for (sequence-specific) transcription factors (TF)[1] (Maston et al. 2006). A unique combination of TFs orchestrates the gene expression in each individual cell. TFs control gene expression by regulating the activity of RNA-Polymerase at their target genes; however, a wide range of additional proteins contribute to this regulation (Figure 1.1). These so-called general transcription factors (GTF) form the pre-initiation complex, nucleosome remodeling enzymes make the DNA accessible, and histone modifying enzymes together with other non-DNA-binding proteins change the local chromatin state to increase or reduce the rate of transcription (Lee and Young 2000; Kadonaga 2004; Maston et al. 2006; Spitz and Furlong 2012). Most of these factors are found at every expressed gene. However, the cell type specific set of genes to which this transcriptional machinery is recruited is directed by sequence-specific TFs, with each TF having a particular set of target genes. To regulate their target genes, TFs bind specific DNA-elements, so-called *cis*-regulatory elements (CRE); which are DNA sequences of several hundred basepairs in length that contain binding sites for multiple TFs (Lelli et al. 2012; Spitz and Furlong 2012).

CREs were initially considered to be located mainly in the vicinity of a gene's transcription start site (TSS) (Lee and Young 2000; Lenhard et al. 2012). This assumption originated from studies in bacteria, where the regulatory information is located mainly in promoters directly adjacent to a gene (reviewed in Haugen et al. 2008). Similarly, in the eukaryotic yeast gene expression is

1 The use of the term transcription factor (TF) varies in the literature. If not stated otherwise, I refer sequence specific transcription factors when using the term TF.

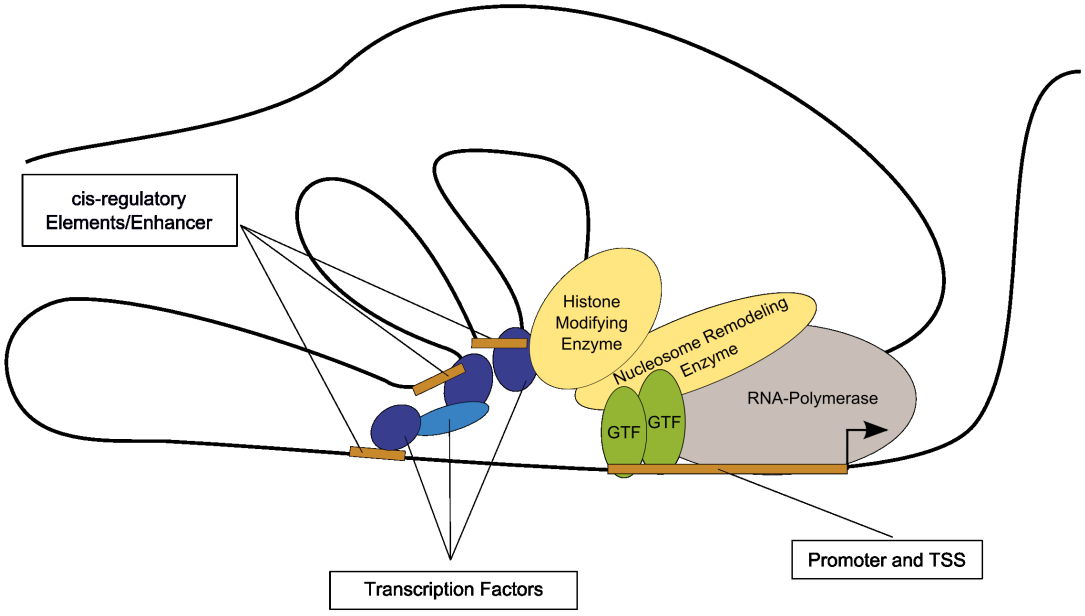


Figure 1.1 General and specific regulatory factors involved in the regulation of gene expression.

Scheme of proteins involved in the control of gene expression. DNA-binding (dark blue) and non-DNA-binding (light blue) TFs, nucleosome remodeling, and histone modifying enzymes facilitate the pre-ininitiation-complex, consisting of RNA-Polymerase and general TFs (GTF) to assemble at a Transcription Start Site (TSS). The specificity in gene regulation is brought about by proximal and distal cis-regulatatory elements (CREs) to which TFs bind.

also mostly controlled at the gene promoter. These findings subsequently led to a large body of research investigating the interplay between TFs and promoters (e.g. Struhl 1981; Hill et al. 1986). In parallel, individual CREs were discovered in the larger genomes of animals, frequently located in intergenic regions and often separated by thousands of basepairs from the genes they control. The widespread distribution of CREs raised the question how TF-binding in the genome is distributed and how TFs identify their target genes (Nobrega et al. 2003; Ruf et al. 2011)? Using the available methods at the time, only investigation with a single or a few CREs or DNA-sequences was feasible. Only in the last decade, major technological advances have enabled new experimental approaches and TF function is now increasingly investigated by profiling the genome-wide binding of TFs in various organisms and cell types (Johnson et al. 2007). The insights gained through these studies radically changed the understanding of TF-DNA interaction: TFs bind to thousands of sites in the genome and most TF binding sites are located in intergenic regions that can be hundreds of kilobases away from the nearest gene (He et al. 2011; Junion et al. 2012; Wang et al. 2012b).

However, these findings created two unresolved questions addressing TF function. First, not each binding site will be of equal importance; therefore, how can functional binding sites be distinguished from non-functional binding sites? Second, if functional binding sites in distant CREs can be located far away from the target genes, how can a TF binding site be associated with the regulated gene?

1.1.1 Transcription Factor Target Recognition

TF proteins bind DNA via specialized DNA-binding domains. The DNA-binding domain usually recognizes a short (6-12bp), degenerate DNA sequence: its recognition sequence (reviewed in Luscombe et al. 2000). A common biochemical mechanism of DNA-binding domains is an alpha helix that anneals into the major grove of the DNA, which allows the amino-acid side chains to directly contact the DNA-base-pairs (Figure 1.2A). Thereby, depending on the amino-acid composition of the alpha helix, the protein can distinguish between different DNA sequences. Unlike restriction enzymes, the sequence-specificity of TFs is often variable and allows a number of similar sequences to be recognized. For example, the preferred binding site of a TF could be the sequence TATTAGC; but it might also bind to TAATAGC or TATTACC sequences, although with lower affinity. But if there are too many variations from the optimal sequence (for example TAATCGA) or when crucial residues in the recognition sequence are altered, the TF will not bind. To graphically represent this variability, sequence logos representing DNA-motifs where developed (Figure 1.2B) (Schneider and Stephens 1990). The size of each letter used in these models signifies the frequency with which each base occurs in a set of binding sites (D'Haeseleer 2006).

Additionally, TFs can change their recognition sequence when binding together with other DNA-binding TFs (Figure 1.3). For instance, in the absence of cofactors the *Drosophila* PBX-ortholog Extradenticle (Exd) binds TTTGACA and the HOX1-ortholog Labial (Lb) preferably binds the sequence TAATTA. However, the two proteins can form a heterodimer that binds a TGATTGATGG sequence, where Exd recognizes the TGAT bases and Lb the TGATGG of the composite recognition sequence (Joshi et al. 2007; Mann et al. 2009).

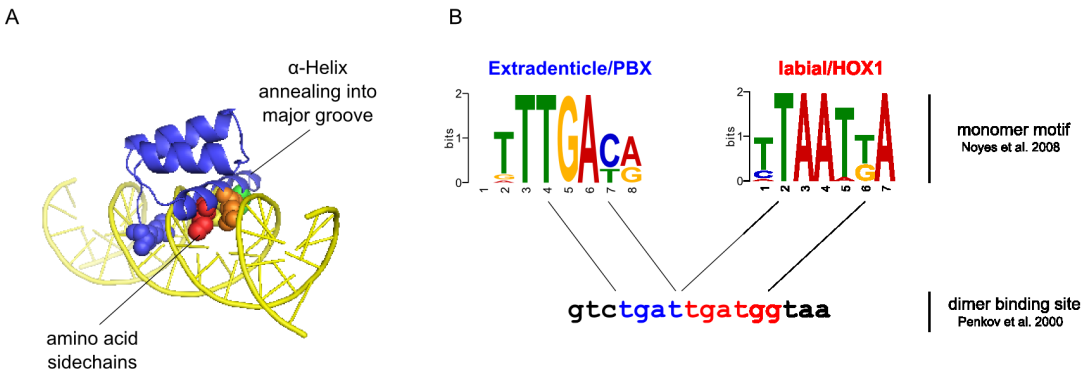


Figure 1.2 Transcription factor DNA recognition.

(A) A homeodomain (*Drosophila* Engrailed (Fraenkel et al. 1998)) annealed to DNA with the amino-acids conferring sequence-specificity highlighted. (B) Sequence logos of Extradenticle (Exd) and Labial (Lb) and a composite binding site of the dimer.

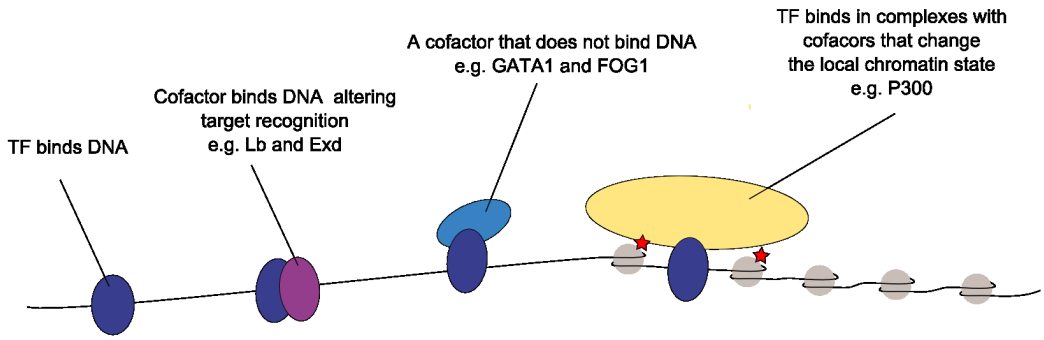


Figure 1.3 Types of TF binding with different co-factors

TFs bind to DNA either alone or in combination with other proteins. The different types of cofactors mediate TF effects e.g. altering the recognition site or changing the local chromatin

Such a mechanism ensures that certain binding sites will only be recognized if both TFs are present and, with hundreds of TFs in the genome, allows DNA-binding of TFs to be both variable and highly specific.

1.1.2 Cofactors

Transcription factors control gene expression by recruiting the transcriptional machinery to their target genes. Although the specific biochemical mechanism is not fully understood, it is evident that TFs are incorporated in large multi-protein complexes that mediate this process (Lenhard et al. 2012). Generally speaking, the term “cofactor” can be applied to any protein in those complexes. More stringently a cofactor is a protein that directly interacts directly with the TF and, depending on type of protein, can lead to different effects on the DNA-binding TF (Figure 1.3). TFs bind cofactors via protein-interaction domains that are separate from the DNA-binding domain. In other cases, such as for Lb and Exd, the DNA-binding domain also mediates the cofactor binding.

Functionally, cofactors are thought to bring about versatility to TF through site-specific interactions. In fact, many TFs are described as activators for some target genes and repressors for other target genes, which could be due to a modulating effect of a cofactor. In contrast to the Lb-Exd interaction, cofactors can also be non-DNA-binding proteins that bind a TF at a subset of its binding sites. One such protein is Friend of GATA1 (FOG1), a cofactor of the TF GATA-binding factor 1 (GATA1). FOG1 is essential for activation and repression of GATA1 target genes in megakaryocytes and erythrocytes (Tsang et al. 1997). Since FOG1 is a non-DNA-binding protein, the interaction of FOG1-GATA1 at individual loci can be identified by the presence or absence of a GATA1 recognition sequences. In addition, comparison of all FOG1 binding sites might reveal the recognition sequence of novel DNA-binding TF interacting with FOG1.

Another important group of cofactors are histone-modifying enzymes such as histone acetyl-transferases (HAT) or histone deacetylases (HDAC). The acetylation of histones at a CRE or a promoter is needed for transcription, whereas deacetylation through HDACs causes transcriptional repression. In developing *Drosophila* embryos, the TF Dorsal activates its target genes by recruiting the HAT CBP/p300 to many of its binding sites. In absence of Dorsal, CBP/p300 binding at these loci, as well as transcription of the target genes is lost (Holmqvist et al. 2012) .

Taken together, not only the TF binding sites are crucial to TF function, but also the combination of proteins that are cobinding at individual sites. Therefore, careful analysis of TF binding sites can shed light on possible cofactor interaction and thereby molecular mechanisms of TF function.

1.1.3 Homeodomain Transcription Factors

Homeodomain TFs are a group of evolutionary highly conserved TFs that share their DNA-binding domain. The eponymous feature, the homeodomain, was named after the homeotic genes (*HOX* genes), a group of TFs that were discovered through their essential role in anterior-posterior patterning of the *Drosophila* embryo (Lewis 1978). Strikingly, when the vertebrate orthologs were identified, the genes were also found to pattern the vertebrate embryo along the anterior-posterior axis (Akam 1989), indicating that patterning mechanism and gene function is conserved in all animals. Aside from the *HOX* genes, many other homeodomain TFs are pivotal in embryonic development. For example, homeodomain-TFs are required for the formation of the heart (*NKX2-5* (Komuro and Izumo 1993)), the eye (*PAX6* (Ton et al. 1991)), the kidney (*PAX2*, *HOXA/D11* (Gong et al. 2007)), the limb (e.g. genes of the *HOXA*/*HOXD* cluster, *MSX2*, or *PITX1* (Logan and Tabin 1999; Zakany and Duboule 2007)), and most other body parts.

The DNA-binding homeodomain is a 60aa protein domain that contains three alpha helices separated by turns (Figure 1.2A and Figure 1.4). The first and second helices stabilize the domain structure, the third helix – the recognition helix – anneals into the major grove of the DNA and confers sequence specificity. The sequence specificity is mainly mediated by the amino acids at position 47, 50, 51 and 54 of the homeodomain, all located in the recognition helix (Gehring et al. 1994). Additional specificity is conferred by amino acids in the N-terminal arm that frequently contact residues in the minor grove (Berger et al. 2008). Compared to other TF classes, homeodomains have broad, usually AT-rich binding sites. *HOX*-gene homeodomains bind either to a [T/C]ATTA (HOX1-8) or to a TAAAA core sequence (HOX9-13). The individual recognition sequence of a given HOX-TF then varies in the bases preceding the core sequence (Jolma et al. 2013).

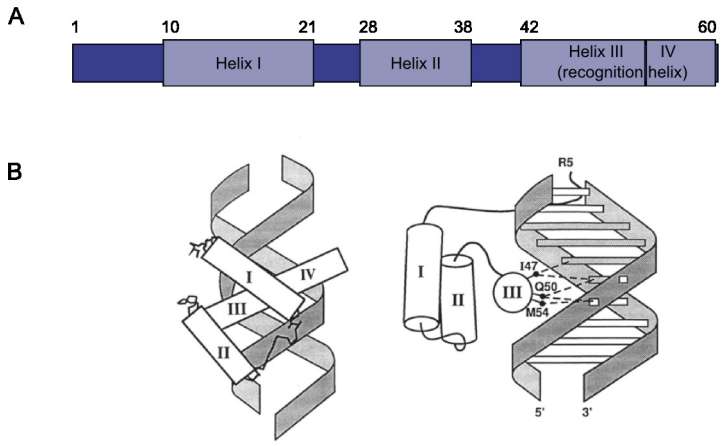


Figure 1.4 Structure of the DNA-binding homeodomain

(A) The 60 amino acid homeodomain and position of the three alpha-helices. (B) Model of a homeodomain bound to DNA (Gehring et al. 1994). Helix III and IV are commonly referred to as helix III/recognition helix.

1.1.4 Transcription Factor Mutations

A mutation in a TF does not only affect the TF protein, but often also influences the expression of its target genes. Since correct gene regulation is of central importance in embryonic development, mutations in human TFs are more prone to cause hereditary diseases than mutations in many other genes (Boyadjiev and Jabs 2000; Vaquerizas et al. 2009). Mutations can cause complex phenotypes affecting seemingly unrelated organs, because many TFs are involved in the development of multiple organ systems and cell types (e.g. the Ulnar-Mammary-Syndrome, which is caused by mutations in *TBX3* (Bamshad et al. 1999)).

Importantly, analysis of TF mutations must consider that each mutation could alter the TF in a specific way. TF mutations in a protein-binding domain may specifically disrupt the interaction with one or more cofactors without affecting the DNA-binding of the protein. Thereby only in a certain cell type, where the cofactor interaction is essential, might be affected, whereas in another cell type that does not required cofactor interaction the TF might function normally (Chlon et al. 2012).

On the other hand, mutations in the DNA-binding domain affect TF function in different ways. A mutation can, for example, completely abolish the DNA-binding of the protein by disrupting the structure of the domain (Todeschini et al. 2011). In consequence, the mutant TF could either be non-functional and act like a complete loss of the protein or might only recognize targets that are bound indirectly via a cofactor (Stender et al. 2010). Furthermore, subtle mutations can lead to similarly drastic consequences. Missense mutations in homeodomain and zinc-finger TFs have been shown to specifically alter the recognition sequence of the TF (Percival-Smith et al. 1990b; Bulyk et al. 2001; Caronia et al. 2003). Thereby, such a mutant TF will remain a fully functional DNA-binding protein, but likely loose many of its original binding sites. Moreover, it could gain new set of binding sites and potentially adapt and misregulate novel target genes.

Therefore, the effects to DNA-binding and protein-binding functions of the mutant TFs need to be carefully examined. In return, these mutations can be instrumental in elucidating the molecular and regulatory function of the TF.

However, the standard approaches currently used for characterization of TF mutations are of general nature and do not take all of these molecular mechanisms into account. In most cases, nuclear localization of the protein is tested in transient transfection assays, which is then supplemented with *in vitro* DNA-binding and/or luciferase assays using reporter constructs. Failure to bind the consensus motif or to activate the reporter results in the characterization of the mutation as a loss-of-function. These experiments mostly allow a simple yes-or-no answer and do not adequately address the complex effects of missense mutations; therefore novel experimental approaches are needed.

1.2 ChIP-seq

For many years, *in vivo* studies of TF-DNA interaction were limited to single or a few loci using Chromatin Immunoprecipitation (ChIP). The development of massively parallel sequencing dramatically changed this, since, combined with new sequencing methods, the ChIP technique enables genome-wide identification of protein-DNA interactions (Johnson et al. 2007).

In the ChIP-seq procedure (Figure 1.5), the chromatin is crosslinked *in vivo* through formaldehyde fixation. Following chromatin purification, protein-DNA complexes are fragmented into short stretches of 200 to 500bp of DNA with the bound proteins crosslinked to it. The fragmented chromatin is then subjected to immunoprecipitation using an antibody directed against the target protein (e.g. a TF). Through this step, the DNA fragments bound by the protein of interest are coenriched and after the immunoprecipitation step the coprecipitated DNA is isolated and sequenced. When mapped to a reference genome, the ChIP-seq method generates a coverage profile in which regions of high coverage (peaks) represent the location of protein-DNA interaction (Figure 1.5). ChIP-seq experiments for various TFs in dozens of cell types have shown that TFs typically bind to thousands of sites in the genome. While this poses enormous challenges to interpreting individual binding sites, it also allows for powerful analysis of TF binding in its *in vivo* environment.

Bioinformatic research groups have developed a plethora of analysis tools that enable versatile evaluation of the enriched regions and address various aspects of TF function. For example, *de novo* motif analysis can identify a TF recognition sequence without prior information. Since TFs bind DNA in a sequence specific manner, the bound genomic sequences will contain the TFs recognition site. By comparing the sequences of all TF binding sites, the recognition sequence can be identified from the ChIP-seq dataset. Furthermore, this analysis can also identify possible cofactors since not only the TFs recognition sequence, but also the recognition sequences of DNA-binding cofactors will be present in the TF binding sites. Unknown cofactor interactions can be identified in ChIP-seq experiments by targeting a non-DNA-binding cofactor. Analysis of the detected binding sites will offer information about the DNA-binding partners of that factor.

Other tools address the genomic location of TF binding sites and comparative analysis of different ChIP-seq datasets. Analysis of ChIP-seq datasets for multiple TFs with related biological function bind has successfully identified CREs by selecting those binding sites, where multiple TFs bind (Junion et al. 2012). Independently, ChIP-seq experiments profiling the genomic distribution of histone modifications achieved to identify CREs independent of TF binding (Heintzman et al. 2007; Ernst and Kellis 2012). Thus, ChIP-seq experiments generate valuable datasets that, when carefully analyzed, offer important insights into TF function and gene regulation.

To date, ChIP-seq has mainly been used to investigate the binding of wildtype TFs. Given the multitude of ways in which a mutation can affect TF function, ChIP-seq offers to be a promising technique for investigating mutant TFs. However, current ChIP-seq methodologies are not well suited to investigate the effects of mutations in TFs, in part because antibodies need to be of high quality and typically do not distinguish between wildtype and mutant proteins. Additionally, TFs often belong to gene families that consist of several close paralogs, which in many cases are coexpressed. A further hindrance is that TFs in developmental processes are active at specific developmental time points for which only very small amounts of tissue would be available in mouse models and which are difficult to recapitulate in cell culture systems. However, if these problems can be overcome, ChIP-seq can be a promising technique to investigate TF mutations.

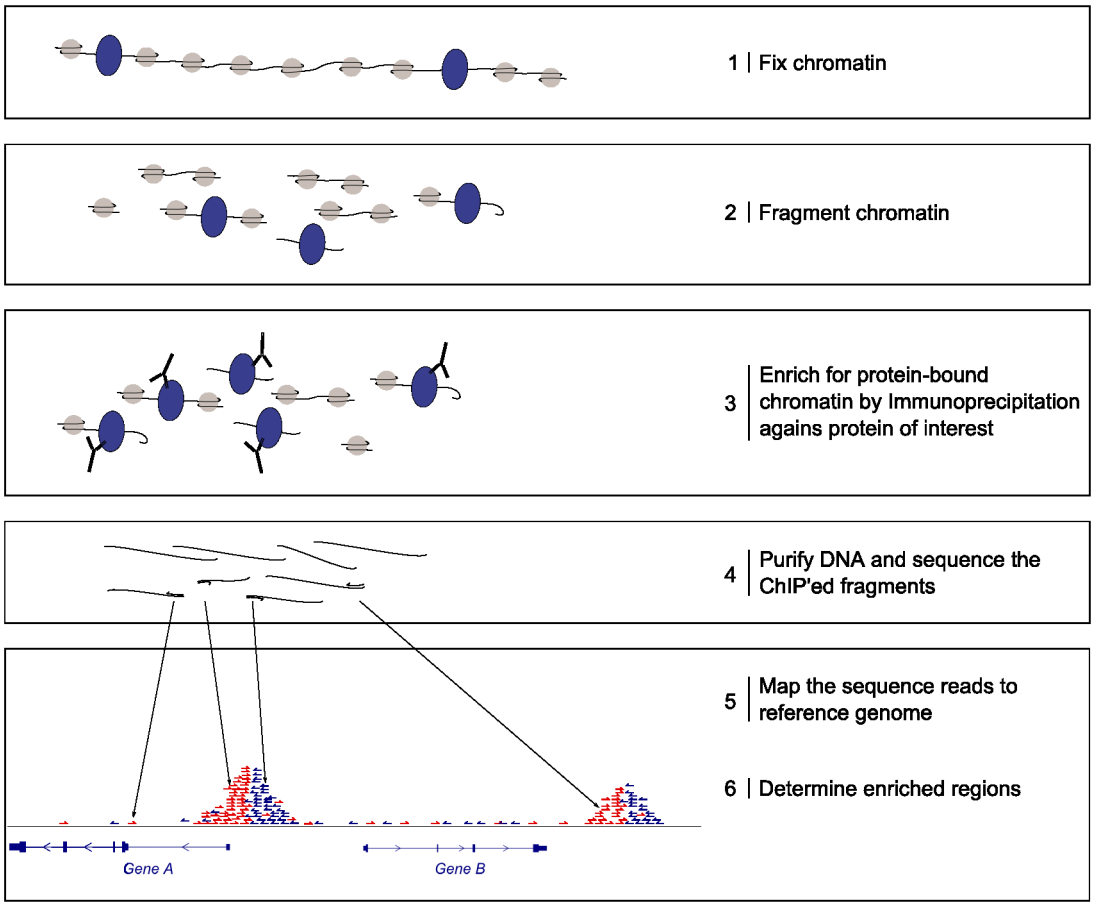


Figure 1.5 Scheme of the ChIP-seq procedure.

Proteins and DNA of the chromatin are crosslinked *in vivo* and subsequently purified from the sample. The chromatin is then fragmented and used for immunoprecipitation with highly specific antibodies, which recognize e.g. TFs bound to DNA or of covalently modified histones. Subsequently, the DNA is purified from immunoprecipitated protein-DNA complexes and sequenced by massive parallel sequencing. When the sequence reads are mapped to a reference genome, protein-bound regions show high coverage with a characteristic, peak-like shape.

1.3 Vertebrate Limb Development

The vertebrate limbs begin to emerge after the primary axis has formed and while axis elongation is still taking place. On each side of the embryo, mesenchymal cells of the lateral plate mesoderm accumulate beneath the overlying ectoderm to form a protrusion, the limb bud. Along the body axis, two pairs of limb buds form and develop into fore- and hindlimbs. The initial limb bud consists of mesodermal mesenchyme that is surrounded by an ectodermal epithelium. These mesenchymal cells will later give rise to the skeleton, connective tissue and tendons of the extremity, while the cells that will form all other organ systems in the limb (muscles, nerves, vascular system, etc.) will invade the growing limb bud at later stages (see review by Tickle 2005).

The undifferentiated early limb bud is already patterned along the three main axes of the extremity, the proximo-distal (shoulder to digit), the dorso-ventral (back-of-the-hand to palm) and the anterior-posterior axis (thumb to pinkie) (Figure 1.6A). Coordinated growth along the three axes then forms the final limb with the stylo-, zeugo-, and autopod; upper extremity, middle extremity, and hands/feet, respectively. The main proliferation occurs in a group of rapidly dividing cells in the so-called progress zone of the distal limb bud (Summerbell 1974; Dudley et al. 2002). The cells of the progress zone receive the proliferative signal from the apical ectodermal ridge (AER), a thickening in the ectoderm along the dorso-ventral midline of the limb bud. In parallel, a region at the posterior end of the limb bud establishes the anterior-posterior axis. This region, the zone of polarizing activity (ZPA), consists of a group of cells that secrete the signaling molecule Sonic hedgehog (SHH), which creates a SHH concentration gradient that determines digit number and identity (Saunders 1948; Riddle et al. 1993). Whereas the interplay between the two signaling centers (AER and ZPA) is well understood, it remains unclear how the signaling gradient is translated into the formation of skeletal elements along the developing limb. Genetic analyses of human mutations and mouse models point to the HOXA and HOXD TFs, that possibly integrate the anterior-posterior and proximo-distal signals originating from the AER and ZPA (Zakany et al. 2004; Zakany and Duboule 2007).

1.3.1 Transcription Factors in Limb Development

Although signaling molecules are the best-described factors governing limb patterning, a myriad of TFs are expressed in the limb bud and show distinctive expression patterns. In combination with the downstream effectors of signaling pathways, TFs pattern the limb bud by creating regionalized gene expression signatures that define the proximal and distal, digital and interdigital regions of the limb (Vokes et al. 2008). Consequently, mutations in TFs have been described to cause distinct limb malformations that affect only specific limb structures. For example, limb malformations of Ulnar-mammary syndrome and Holt-Oram syndrome are

dysplasias of the stylo- and zeugopod and are caused by mutations in the TFs *TBX3* and *TBX5*, respectively (Bamshad et al. 1997; Li et al. 1997) (Figure 1.6B). Townes-Brocks syndrome includes a autopod-restricted polydactyly phenotype and is caused by mutations in the *SALL1* TF (Kohlhase et al. 1998) (Figure 1.6B). Similarly, mutations in the TFs *HOXD13* and *HOXA13* lead to malformations of only the hands and feet (Muragaki et al. 1996; Mortlock and Innis 1997) (Figure 1.6C).

Even though TF expression patterns and knockout models attest to the role of TFs in limb development, the gene regulatory network that underlies limb patterning remains elusive. To highlight the fundamental influence TFs exert on limb development, two of the best-described examples are presented: forelimb-hindlimb specification through *PITX1* and *TBX4/5*, and anterior-posterior and proximo-distal patterning by the *HOXA* and *HOXD* genes.

Determination of Hindlimb Identity by PITX1

The vertebrate fore- and hindlimbs develop as serially homologous structures and share the same general anatomy. Correspondingly, expression patterns of most genes are identical between both limb types (Shou et al. 2005; Duboc and Logan 2011). Still, the final shapes of fore- and hindlimbs are strikingly different from one another, if one considers bats, mice, birds, or humans, for example.

Only few genes are known to be differentially expressed between fore- and hindlimbs (Logan et al. 1998). One of those is the homeodomain transcription factor *PITX1*. In the developing embryo, *PITX1* is expressed throughout the hindlimb bud, but not in the forelimb (Figure 1.6D) (Logan and Tabin 1999; Szeto et al. 1999). Loss of *Pitx1* in murine hindlimb buds causes the hindlimb to develop forelimb-like characteristics (Lanctot et al. 1999). Conversely, gain-of-function experiments that misexpress *PITX1/Pitx1* in the developing forelimb in chicken or mouse embryos cause the forelimb to develop hindlimb-like characteristics, such as the formation of knee-like joints instead of elbows (Logan and Tabin 1999; Szeto et al. 1999). The dramatic change in limb morphology results from misexpression of only one TF. Yet, the targets through which *Pitx1* mediates these effects remain largely unknown. It is believed that regulatory genes, such as other TFs, signaling molecules, or their receptors are among the most crucial target genes. However, identification of key target genes for *Pitx1* has so far not been very successful. Only one hindlimb-specific target has been validated unequivocally, the TF *Tbx4* (Logan and Tabin 1999; Menke et al. 2008; Infante et al. 2012). In an attempt to further elucidate gene regulation by *Pitx1,* a recent study reported the genome-wide binding of PITX1 in murine hindlimbs using ChIP-seq (Infante et al. 2012). The outcome of this study demonstrates the challenges that are a consequence of genome-wide binding profiles. Around 10,000 binding sites for PITX1 were reported, partly overlapping with hindlimb-specific CREs. Reaffirmingly, PITX1 was shown to bind to a number of CREs that control *Tbx4* expression.

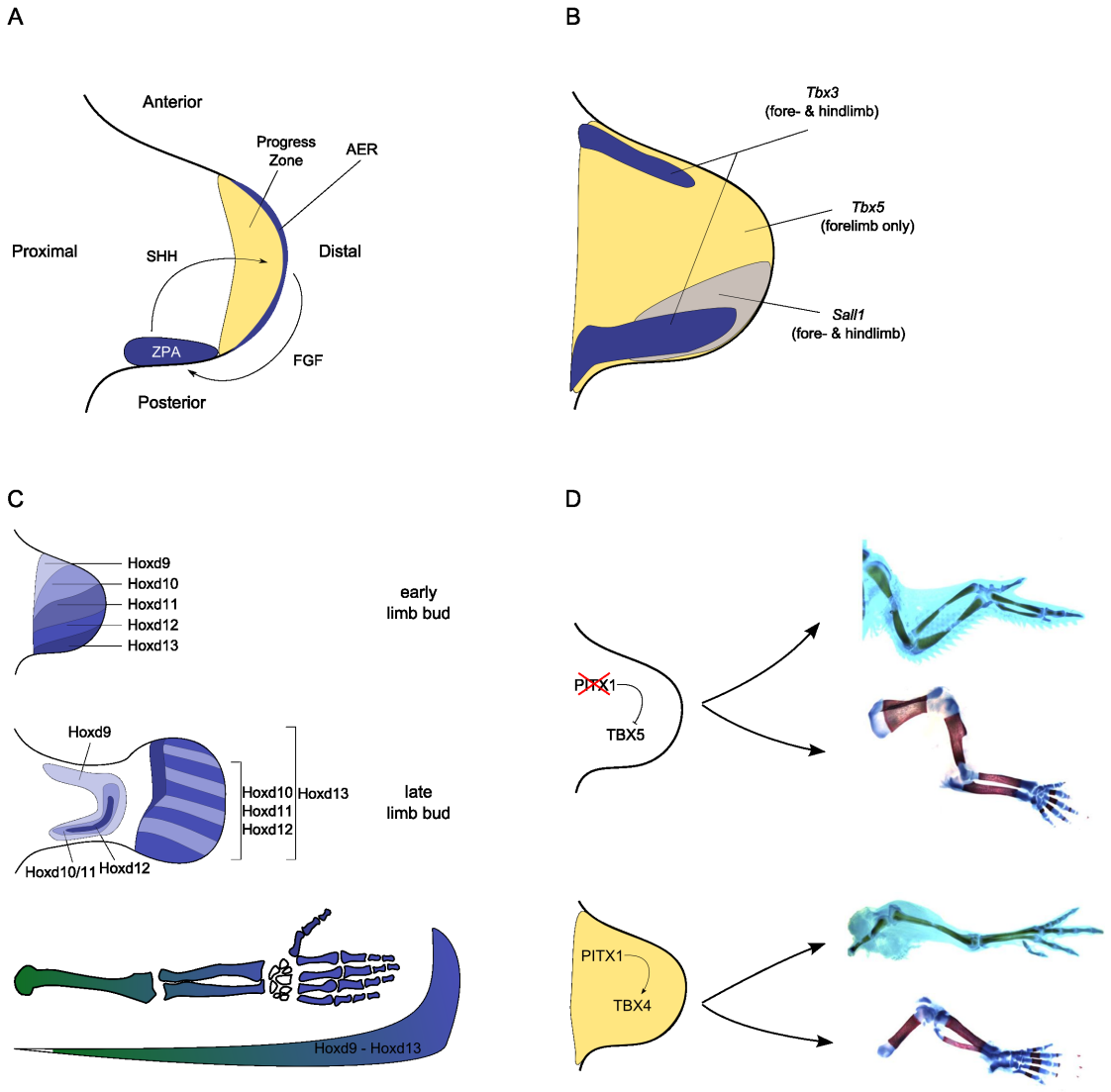


Figure 1.6 Transcription Factors and Limb Development

(A) The two main signaling centers in the limb bud, AER and ZPA. Shh and FGF signalling molecules create a positive feedback loop that drives limb bud outgrowth (modified after Zeller et al 2009). (B) expression pattern of three TFs that cause limb malformations in patients (C) HoxD TFs and their overlapping expression patterns in early and late limb buds. Colinear expression of the genes patterns the limb along the proximodistal and anterioposterior axis (modified after Zakany and Duboule 2007, Woltering and Duboule 2010) (D) Pitx1 determines limb identity in all vertebrates. The Tbx4/5 genes are downstream of Pitx1.

However, just one novel target gene was reported, *Tbx2*, which is expressed in fore- and hindlimbs and does not seem to be directly related to limb identity. This example demonstrates that mere *in vivo* profiling of PITX1 binding sites did not improve the molecular understanding of PITX1 function in determining hindlimb identity.

HOX Genes in Limb Development

Throughout the animal kingdom, *HOX* genes pattern the anterior-posterior axis of the embryo (Lewis 1978; Akam 1989). The most striking feature of this gene family is the interrelation of the genes' genomic organization with their respective expression domains along the main body

axis, a mode of expression termed collinearity (Dolle et al. 1989; Graham et al. 1989; Tarchini and Duboule 2006). *HOX* genes are organized in clusters, and the position of a gene within the cluster correlates with its expression domain along the main body axis; the most 3' gene of the cluster has the most anterior and the most 5' the most posterior expression domain (Carroll 1995).

In tetrapods, there are four *HOX*-clusters, *HOXA* to *HOXD*, comprised of 39 genes in 13 paralog groups[2] (Gehring et al. 2009). The *HOXA* and *HOXD* genes not only pattern the main body axis but also are critical for pattering the limb. Intriguingly, the collinear mode of expression has also been adopted in the limb. The *HOX* genes 3' in the cluster are expressed in the proximal parts of the limb, while the more 5' *HOX* genes have increasingly distal expression domains (Figure 1.6C) (Tschopp et al. 2009; Woltering and Duboule 2010). Analysis of a series of mice carrying targeted deletion of some of the *HOX* genes uncovered their regional influence on limb development. For example, mice missing *Hoxa9*/*Hoxd9* lack the humerus and femur, whereas targeted deletion of *Hoxa10*/*Hoxd10* affects tibia and fibula, and deletion of *Hoxa11*/*Hoxd11* leads to a strong reduction of radius and ulna. Furthermore, similar mouse models deleting *Hoxd11-13* and/or *Hoxa13* demonstrated that these genes control digit development (reviewed in Zakany and Duboule 2007). In addition to regional pattering, the mutations also affect the size and shape of the particular bones (Woltering and Duboule 2010).

In early stage limb buds, the posterior *HOX* genes act redundantly to activate *Shh* expression in the ZPA (Zakany et al. 2004; Kmita et al. 2005). At a later stage, when the digits form, *Hoxd13* has been shown to mediate retinoic acid induced interdigital cell death (Kuss et al. 2009). Moreover, *Hoxd13* and *Hoxa13* both influence skeletal development (Salsi et al. 2008; Kuss et al. 2009; Perez et al. 2010; Villavicencio-Lorini et al. 2010).

Despite major progress in understanding the regulation of the *HOX* gene clusters in limb development, only little is known about the targets of those TFs (Svingen and Tonissen 2006). Limb pattering and skeletal development suggest a complex mixture of HOX target genes that could on the one hand include genes of signaling pathways involved in patterning and on the other hand genes regulating specific skeletal cell differentiation processes. In line with these findings, human mutations in *HOX* genes are associated with digit malformations, which can affect both, limb pattering (i.e., digit number) and skeletal development.

[2] *HOXA1-13*, *HOXB1-13*, *HOXC1-13*, and *HOXD1-13*. Each cluster misses some of the paralogs (e.g. there is no *HOXA12*)

1.4 Pathogenic Mutations in *HOXD13*

Mutations in *HOXD13* were the first mutations identified in a human *HOX* gene (Muragaki et al. 1996). The phenotype most frequently associated with *HOXD13* mutations is synpolydactyly (SPD, MIM *186000). SPD is characterized by a syndactyly of the third and fourth fingers and the fourth and fifth toes, with a partial or complete digit forming in the syndactylous web. However, the phenotypic spectrum of *HOXD13* mutations encompasses various other phenotypes, some of which are described below.

More than 20 distinct mutations *HOXD13* have been reported (Figure 1.7A) (for a recent review, see Brison et al. 2013). The mutations are distributed along the whole gene body, although there is an accumulation of mutations that alter either the N-terminal polyalanine stretch or the C-terminal homeodomain (Figure 1.7).

Each mutation can be assigned to one of three classes:

I) Null Mutations
II) Polyalanine Tract Variations
III) Missense Mutations

Class I – Null Mutations

Four nonsense (Furniss et al. 2009; Kurban et al. 2011; Brison et al. 2012b; Jamsheer et al. 2012) and three frameshift (Goodman et al. 1998; Brison et al. 2012b) mutations have been reported, all associated with a SPD phenotype. In the heterozygous state the SPD has a weak penetrance (around 30%) and expressivity but is fully penetrant in the homozygous state (Goodman et al. 1998; Kan et al. 2003). All alleles are thought to be null alleles due to nonsense mediated mRNA decay, although no experimental validation has been reported. Thus, the heterozygous phenotype would result from functional haploinsufficiency. Similar to the mild effects of haploinsufficiency in humans, a *Hoxd13* mouse knockout model shows no obvious limb phenotype in heterozygous state but a strong phenotype when both copies of *Hoxd13* are missing (Dollé et al. 1993).

Class II – Polyalanine Tract Variations

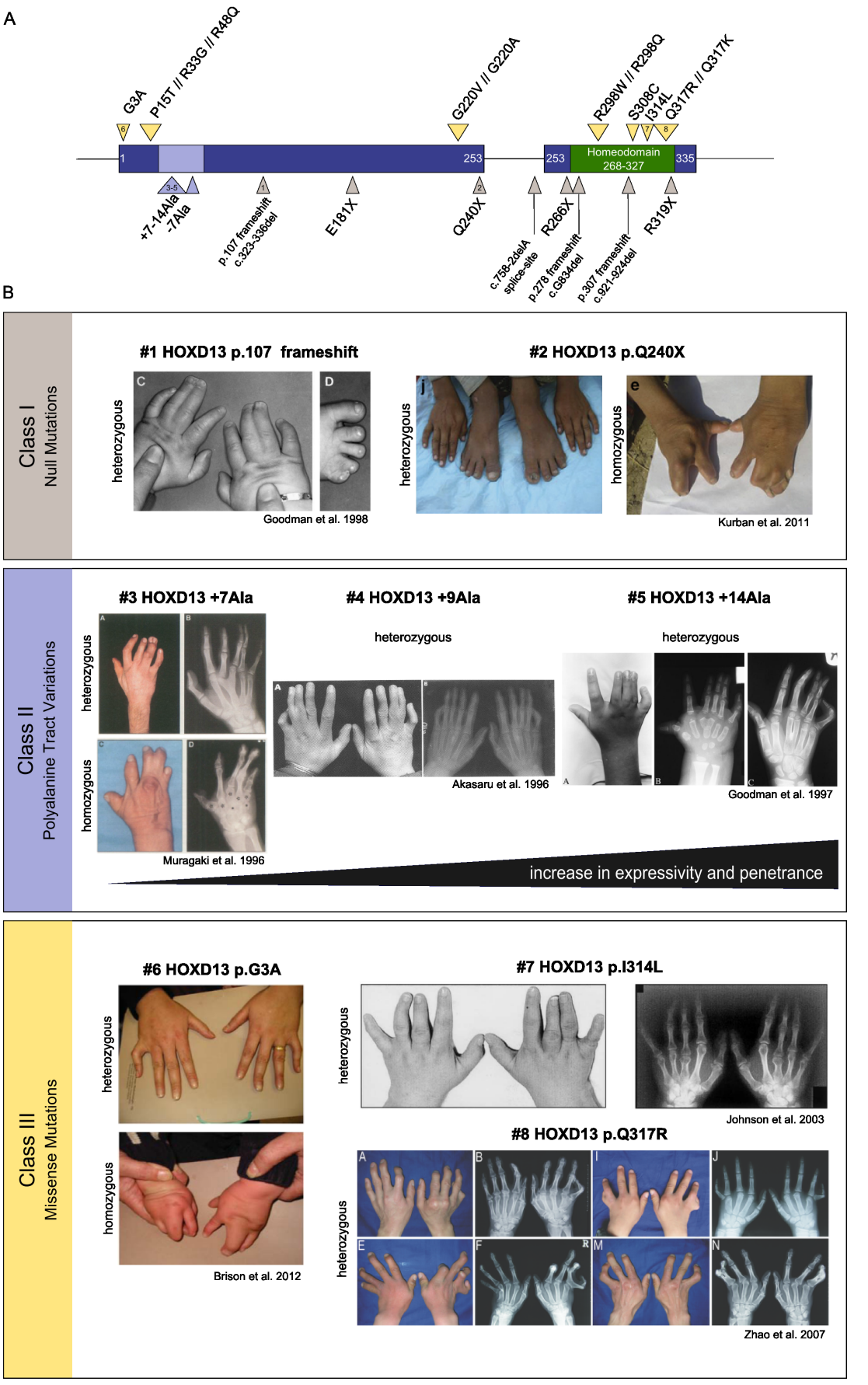
Near the N-terminus, HOXD13 harbors a trinucleotide repeat encoding a stretch of 15 alanines in the protein (Figure 1.7). Expansions of the polyalanine tract were the first described *HOXD13* mutations and expansions of seven to 14 alanines have been reported in over 40 families with a SPD phenotype (Brison et al. 2013). Strikingly, SPD penetrance and expressivity increase with the length of the alanine expansion (Goodman et al. 1997). Further, homozygous individuals are

more severely affected than heterozygous and show an additional phenotype characterized by malformation of the metacarpal bones, which has been interpreted as a metacarpal-to-carpal homeotic transformation (Kuss et al. 2009).

Functional analysis of HOXD13^{polyAla} proteins demonstrated that the expanded polyalanine tract of HOXD13 causes aggregation of the protein, its retention in the cytoplasm, and finally its degradation (Albrecht et al. 2004). In addition, the mutant copy of HOXD13 acts dominant-negatively by entrapping some of the wildtype proteins, which then also leads to partial retention and degradation of wildtype HOXD13 protein. Thereby the polyalanine expansion does not only affect the mutant protein, but also renders some of the remaining functional HOXD13 inert. Finally, the mutant proteins' cytoplasmatic retention becomes more pronounced the longer the extension of the polyalanine-repeat is, and thus provides a molecular basis for the phenotypic observation described above (Albrecht et al. 2004; Villavicencio-Lorini et al. 2010). Finally, three cases of polyalanine-contractions have been reported in patients with a brachydactyly phenotype in the hands and a SPD-like phenotype in their feet (Nakano et al. 2007; Zhao et al. 2007; Garcia-Barceló et al. 2008), but there is currently no functional explanation for the underlying pathomechanism.

Figure 1.7 Pathogenic mutations in HOXD13 are associated with several phenotypes (next page).

(A) Scheme of the *HOXD13* gene locus and annotation of published human mutations in HOXD13. The missense mutations are indicated by yellow arrowheads, polyalanine tract variations in blue and putative null mutations in gray. The numbers refer to photographs in (B). The three classes of HOXD13 mutations are associated with distinct dysmorphologies, i.e. synpolydactylies and different types of brachydactylies.


A
G3A
P15T // R33G // R48Q
G220V // G220A
R298W // R298Q
S308C
I314L
Q317R // Q317K
1
253
Homeodomain
268-327
335
+7-14Ala
-7Ala
p.107 frameshift
c.323-336del
E181X
Q240X
c.758-2delA
splice-site
p.278 frameshift
c.G834del
p.307 frameshift
c.921-924del
R266X
R319X
B
Class I
Null Mutations
#1 HOXD13 p.107 frameshift
heterozygous
Goodman et al. 1998
#2 HOXD13 p.Q240X
heterozygous
homozygous
Kurban et al. 2011
Class II
Polyalanine Tract Variations
#3 HOXD13 +7Ala
heterozygous
homozygous
Muragaki et al. 1996
#4 HOXD13 +9Ala
heterozygous
Akasaru et al. 1996
#5 HOXD13 +14Ala
heterozygous
Goodman et al. 1997
increase in expressivity and penetrance
Class III
Missense Mutations
#6 HOXD13 p.G3A
heterozygous
homozygous
Brison et al. 2012
#7 HOXD13 p.I314L
heterozygous
Johnson et al. 2003
#8 HOXD13 p.Q317R
heterozygous
Zhao et al. 2007

Class III – Missense Mutations

To date, twelve missense mutations in *HOXD13* have been reported, six of which are mutations in the DNA-binding homeodomain. Three missense mutations outside the homeodomain (G3A, G220A, and G220V) cause a SPD phenotype and additional clinodactyly of the 5th finger with varying degrees of expressivity and penetrance, not unlike class I mutations (Fantini et al. 2009; Brison et al. 2012a; Zhou et al. 2013)[3]. In contrast, missense mutations in the homeodomain cause multiple phenotypes that can include SPD-like dysplasias but can also appear completely unrelated.

A R298W substitution with reduced penetrance is reported in a Belgian family with 17 mutation carriers; 13/17 heterozygous mutation carriers display bilateral clinodactyly of the 5th finger but a stronger effect is seen in three patients, which display a unilateral SPD phenotype (Debeer et al. 2002). A second dominantly inherited missense mutation affecting the same residue, R298Q, is reported in a Chinese family (Wang et al. 2012a). The mutation causes a fully penetrant SPD. Additionally, 3/6 individuals show a brachydactyly A2 phenotype (BDA2) that has not been reported for any other *HOXD13* mutation. This study adds two new cases with an identical *de novo* mutation, leading to a R298Q substitution. Bot patients, like in the previously reported family show bilateral SPD and BDA2.

The remaining three missense mutations in the HOXD13 homeodomain do not cause a SPD phenotype. Two mutations, S308C and I314L, are associated with phenotypes overlapping with brachydactylies type E and D. The phenotype has been reported to overlap marginally with the features of SPD, because isolated patients display a partial polydactyly of the 4th finger (Johnson et al. 2003). Molecularly, the I314L mutation has been shown subtly alter the DNA-binding specificity by preferably binding to only one of two possible HOXD13 recognition sequences (Caronia et al. 2003). Yet another phenotype is caused by a Q317R substitution in a Chinese family with dominantly inherited syndactyly type V (Zhao et al. 2007). The affected Q317 corresponds to position 50 in the homeodomain that, like the I314 (I47), is one of the amino acids in the homeodomain known to directly contact the DNA basepairs and is therefore required for sequence specificity. The mutation is reported to act as a partial loss-of-function. However, this finding relies on a single functional assay that shows reduced activation of a reporter construct by the mutant. Finally, this work describes a *de novo* missense mutation in HOXD13 that was discovered in a patient with a complex brachydactyly/oligodactyly phenotype that has not been reported with any other HOXD13 mutation or with other mutations. The mutation converts Q317 to a lysine (Q317K) and although it affects the same residue as the Q317R substitution the phenotypes do not resemble each other.

[3] The remaining three missense mutations outside the homeodomain (P15T, R33G and R48Q) were identified in a patient screen by Nakano K, Sakai N, Yamazaki Y, Watanabe H, Yamada N, Sezaki K, Susami T, Tokunaga K, Takato T, Uchinuma E. 2007. Novel mutations of the HOXD13 gene in hand and foot malformations. *International surgery* **92**(5): 287. and are reported to cause Wassel type IV polydactyly of the thumbs and SPD in the feet. However, the publication does not show patient photographs to assess the severity of the phenotype.

Given the preliminary evidence, each missense mutation inside the HOXD13 homeodomain seems to lead to individual alterations of protein function, thereby changing affinity and/or specificity of the TF. In turn such specific effects would help to explain the wide phenotypic spectrum of *HOXD13* mutations.

1.5 Aims and Objectives

This thesis aims to further our understanding of the molecular mechanisms underlying missense mutations in TFs. Therefore, two novel pathogenic mutations in *HOXD13* – Q317K and R298Q – were investigated. Both missense mutations affect the DNA-binding homeodomain of HOXD13, but cause very different patient phenotypes. The Q317K mutation is associated with a severe form of oligodactyly combined with brachydactyly, whereas the R298Q mutation is associated with SPD and BDA2.

Current techniques, such as testing the nuclear localization, activation of reporter constructs, or *in vitro* DNA-binding assays provide only limited insights into the molecular changes induced by TF mutations. Given the diverse aspects to TF function, these methodologies are not sufficient to examine differences between different mutations and new experimental approaches are required.

In this study, a new cell-culture based ChIP-seq approach was used to compare wildtype and mutant HOXD13 proteins. Central to this thesis is the application of the ChIP-seq technique, which allows versatile analysis of the results with regard to different aspects of TF function. Genome-wide binding profiles of wildtype and mutant HOXD13 were produced and, following initial data analysis, mutation-specific hypotheses were generated. Subsequently, the hypotheses regarding the molecular pathomechanism were tested using appropriate functional experiments for the Q317K and R298Q mutation.

In combination, the functional analysis of the two HOXD13 mutations based on the ChIP-seq system revealed clearly distinct pathomechanisms for the Q317K and the R298Q missense mutations, and thereby point to different molecular requirements for correct HOXD13 function.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Unless stated otherwise, chemicals were obtained from Merck (Darmstadt), Roth (Karlsruhe) or Sigma-Aldrich (Hamburg, Seelze, Schnelldorf and Steinheim) in analytical grade quality.

2.1.2 Buffers

Common buffers and solutions were prepared according to Sambrook et al. 2001.

Buffers for Chromatin-Immunoprecipitation

Lysis Buffer 1:	50mM HEPES-KOH, pH7.5; 140mM NaCl; 1mM EDTA; 10% Glycerol; 0,5% NP-40; 0,25% Triton X-100; Protease Inhibitors (Roche complete, add fresh)
Lysis Buffer 2:	10mM Tris-HCl, pH 8.0; 200mM NaCl; 1mM EDTA; 0.5mM EGTA; Protease Inhibitors (Roche complete, add fresh)
Lysis Buffer 3:	10mM Tris-HCl, pH 8.0; 100mM NaCl; 1mM EDTA; 0.5mM EGTA; 0,1% Na-Deoxycholate; 0,5% *N*-Laurylsarcosine; Protease Inhibitors (Roche complete, add fresh)
RIPA (Wash Buffer):	50mM HEPES-KOH, pKa 7.55; 500mM LiCl; 1mM EDTA; 1,0% NP-40; 0,7% Na-Deoxycholate; Protease Inhibitors (Roche complete, add fresh)
TE-NaCl:	10mM Tris-HCl, pH 8.0; 1mM EDTA; 50mM NaCl, Protease Inhibitors (Roche complete, add fresh)
ChIP-Elution Buffer:	50 mM Tris-HCl, pH8.0; 10mM EDTA, 1.0% SDS

Buffers for Electromobility Shift Assays

Binding Buffer:	100 mM NaCl; 2 mM $MgCl_2$; 0,1 mg/ml BSA; 4 mM spermidine; 25 mM HEPES, pH7.5; Protease Inhibitors (Roche complete)
Loading Buffer:	(40% Glycerol + 0.01% Bromphenol blue)

2.1.3 Antibodies

Table 2.1 Antibodies and Immuncytochemistry Working Dilutions

Target	Concentration/Dilution	Catalog # / Lot #	Supplier
αFLAG-M2	1mg/ml 1:200-300	F1804-5MG / 035K6196	Sigma-Aldrich, DE
gtαm-ALEXA546	2mg/ml 1:200	A11003 / 34779A	Molecular Probes, NL

Protein G magnetic beads were obtained from Invitrogen (Dynabeads, Cat.-No. #100.04D)

2.1.4 Kits

Standard procedures, such as DNA or RNA purification, cDNA-synthesis, DNA-sequencing, protein concentration measurement were conducted following manufacturer's instructions using listed kits.

Table 2.2 Molecular Biology Kits

Task	Kit	Supplier
Plasmid DNA-purification	NucleoSpin Plasmid	Macherey-Nagel, Düren
Plasmid DNA-purification	Nucleobond PC100	Macherey-Nagel, Düren
Plasmid DNA-purification	Nucleobond PC100 EF	Macherey-Nagel, Düren
DNA-purification	ZymoClean Gel DNA Recovery Kit	Zymo Research, Irvine, USA
RNA-purification	RNeasy Mini Kit	Qiagen, Hilden
cDNA-synthesis	TaqMan Reverse Transcription Reagents	Applied Biosystems, Foster City, USA
Sanger-Sequencing	BigDye Terminator v3.1 Sequencing Kit	Applied Biosystems, Foster City, USA
Protein concentration measurement	BCA Protein Assay Kit	Pierce/Thermo Fisher Scientific, Bonn

2.1.5 Enzymes

Restriction enzymes were obtained from NEB (Frankfurt) or MBI-Fermentas (St. Leon-Roth). Taq- and Pfu-DNA-polymerases were produced in house (A.C. Stiege). Phusion DNA-Polymerase was obtained from NEB, T4-ligase and Polymerase from MBI- Fermentas, and RNase A (Cat.-No. R4875) and Proteinase K (Cat.-No. P2308) from Sigma-Aldrich.

2.1.6 Bacterial Strains

General cloning steps were performed in the *E. coli* Top10 (Invitrogen) strain. The expression strain used for production of *Hoxd13* homeodomains was *E. coli* Rosetta2 (DE3) pLysS (Novagen).

2.1.7 Expression Constructs and Vectors

Table 2.3 Plasmid Vectors used in this study

Plasmid	Task	Supplier
pSLAX13-5'FLAG	Modified RCAS-Shuttle Vector	Dr. Jochen Hecht (Berlin)
RCASBP(A)-5'	Modified RCASBP-Vector	Dr. Jochen Hecht (Berlin)
RCASBP(B)-5'	Modified RCASBP-Vector	Dr. Jochen Hecht (Berlin)
pTA-GFP	Cloning of in situ probes	Dr. Jochen Hecht (Berlin)
pVAX1	Expression vector for luciferase assays	Invitrogen
pCMV-mPitx1	Transient expression vector	Kind gift of Dr. J. Drouin and Dr. J. Johnston
pET-41c	Bacterial expression vector for expression of HOXD13 homeodomains	Millipore

2.1.8 Primer

All primers were synthesized by MWG Biotech AG (Ebersberg) and HPSF purified.

Table 2.4 Primers for absolute quantification of RCAS-derived constructs

Oligoname	Sequence (5' → 3')	
D13splice_F	GAGCTGAGCTGACTCTGCTGGTGG	Cloning RCAS standard plasmid
D13splice_R	CTTAATGACGGCTTCCATGCTTGATC	
D13qRT-F1	TCATCCTTTCCAGGAGATGTGG	measuring all viral RNA
D13qRT-R1	AGGCACCCTTTTCTTGCTTCC	
RCAS-spl-F1	AACTCAGAGGGTCGTCGGAAG	measuring Hoxd13-spliced RNA
RCAS-spl-R2	TCACCGTCATCGTCTTTGTAGTC	
cPITX1_RT_F	AACCGCTACCCCGATATGAG	
cPITX1_RT_R	ACAGGTCCATCTGCTGGTTC	
cHOXD13_RT_F	AGCTCGCACTTCTGGAAATC	

cHOXD13_RT_R	TAAGCTGGAGCTTGGTGTAGG
cGAPDH_RT_F	CCATTCCTCCACCTTTGATG
cGAPDH_RT_R	CACGGTTGCTGTATCCAAAC
cHOXD13_RT_F	AGCTCGCACTTCTGGAAATC
cHOXD13_RT_R	TAAGCTGGAGCTTGGTGTAGG

Table 2.5 Primers used for quantitative Realtime-PCR

Oligoname	Sequence (5' → 3')
cGAPDH_RT_F	CCATTCCTCCACCTTTGATG
cGAPDH_RT_R	CACGGTTGCTGTATCCAAAC
cBMP4_RT_F	TGGGAGACCTTTGATGTGAG
cBMP4_RT_R	TTTGCCCTGATGAGTCTGTG
cFRAS1_RT_F	ATGGCCGACTGGTTATTGAG
cFRAS1_RT_R	TGACTTCACATCTGGCAAGC
cMSX2_RT_F	CGAGAGAGACTTCCTGGTGG
cMSX2_RT_R	CGGTTGGGTACTGCATTCTT

Table 2.6 Primers used in Electromobility-Shift-Assays (EMSA)

Oligoname	Sequence (5' → 3')
Cy-D13-f-5'	CY3-ggatcCCAATAAAAtcggc (labeled oligo)
Cy-D13-f-3'	CY3-ccgaTTTTATTGGgatcc (labeled oligo)
D13-f-5'	ggatcCCAATAAAAtcggc (unlabeled competitor)
D13-f-3'	ccgaTTTTATTGGgatcc (unlabeled competitor)
D13mut-f-5'	ggatcCCcAgcAcAtcgg (mutant unlabeled competitor)
D13mut-f-3'	ccgaTgTgcTgGGgatcc (mutant unlabeled competitor)
Cy-PITX-f-5'	CY3-ggatcAGGGATTAACtcgg (labeled oligo)
Cy-PITX-f-3'	CY3-ccgaGTTAATCCCTgatcc (labeled oligo)
PITX-f-5'	ggatcAGGGATTAACtcgg (unlabeled oligo)
PITX-f-3'	ccgaGTTAATCCCTgatcc (unlabeled oligo)
PITXmut-f-5'	ggatcAGGGcggAACtcgg (mutant unlabeled competitor)

Table 2.7 Primer used for cloning of expression and luciferase vectors

Oligoname	Sequence (5' → 3')
Gg_PITX1_NcoI_F	GCCCCCGCCATGGATTCC
Gg_PITX1_SpeI_R	ACTAGTCCGTCAACTGTTGTACTGAC
Mm_Pitx1_NcoI_F	CACCTCCATGGACGCCTTCAAGGGAGG
Mm_Pitx1_RI_R	AAAGAATTCGGTCAGCTGTTGTACTGGCAAGCGT
Gg_SOX9_F3_Nhe1	AAAGCTAGCAAGAGAAGCGCAACGGC
Gg_SOX9_R1_Xho1	AAAACTCGAGAGTAAAGCCGGCAGAGACG

2.1.9 Cultured Cell Lines

Table 2.8 Cell culture lines

Cell Line	Origin	Supplier
DF-1	Chicken fibroblast	ATCC
NIH3T3	Murine embryonic fibroblasts	ATCC

2.1.10 Animals

Fertilized eggs (Clean Eggs quality) for chicken micromass cultures were obtained from VALO BioMedia GmbH (Osterholz-Scharmbeck).

2.1.11 Instruments

Table 2.9 Centrifuges

Instrument	Model No. / Type	Supplier
Table Top centrifuge	5414D	Eppendorf, Hamburg
Chilling centrifuge	5417R	Eppendorf, Hamburg
Microtiterplate centrifuge	5416	Eppendorf, Hamburg
Chilling centrifuge	Avanti J-E	Beckman-Coulter, Palo Alto, USA
Rotor	JLA16250	Beckman-Coulter, Palo Alto, USA
Very cool refrigerating centrifuge	Megafuge 1.0	Thermo Fisher
Ultracentrifuge	L7-55	Beckman, Palo Alto, USA
Ultracentrifuge Rotor	SW 32-Ti	Beckman, Palo Alto, USA

Table 2.10 Thermocycler

Instrument	Model No. / Type	Supplier
Thermocycler	GeneAmp PCR System 2700, 2720 and 9700	Applied Biosystems, Foster City, USA
Real-time Cycler	ABIPrism 7900 HT	Applied Biosystems, Foster City, USA

Table 2.11 Microscopes

Instrument	Model No. / Type	Supplier
Stereomicroscope	MZ7-5	Leica, Bensheim
Camera	Axiocam MRc5	Zeiss, Göttingen
Light source	KL1500 LCD	Leica, Bensheim
Software	Axiovision 4.x	Zeiss, Göttingen

Table 2.12 Next-Generation Sequencing and other

Instrument	Model No. / Type	Supplier
Sequencer	Genome Analyzer IIX	Illumina, USA
Cluster Station		Illumina, USA
Sonicator	BioRuptor NextGen UCD-300	Diagenode, Belgium
Photometer	NanoDrop 2000	Thermo Scientific, Wilmington, USA
Microplate Reader	Spectra Max 250	Molecular Devices, Biberach
Fluorescent imaging system	FLA-5000	Fuji, Tokyo, Japan
Biochemical Analysis	Bioanalyzer 2100	Agilent Technologies, Böblingen

2.1.12 Software

General Software

Digital pictures were edited using Carl Zeiss Axiovision 4.8.2 and Adobe Photoshop CS4. Figures were composed using Inkscape 0.48. The bibliography was managed using Thomson Reuters Endnote.

Bioinformatics Software

Table 2.13 lists the bioinformatics software used to perform all ChIP-seq analyses.

Table 2.13 Specialized bioinformatics software

Software	Task	Reference
FastQC	Quality Control fastq-sequencing files	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
BWA Aligner	Aligning NGS-sequenced reads to reference genome	(Li and Durbin 2009)
SAM-Tools	Handling of SAM-files	(Li et al. 2009)
MEME-suite	Various motif analysis tools	(Bailey et al. 2009)
SPP	Quality Control of ChIP-enrichment	(Kharchenko et al. 2008)
BED-Tools	Handling of BED-files	(Quinlan and Hall 2010)
MACS2	Detection of ChIP-enriched regions	(Zhang et al. 2008)
DREME	Motif-analysis	(Bailey 2011)
IDR	Reproducibility of ChIP-seq experiments	(Li et al. 2011; Landt et al. 2012)
seqMINER	Read distribution analysis of ChIP-seq datasets	(Ye et al. 2011)

2.2 Methods

2.2.1 General Molecular Biological Methods

All standard molecular biological procedures were performed according to Sambrook et al 2001.

DNA Isolation

Isolation of plasmid-DNA from *E. coli* was performed according to manufacturer's instructions using the Nucleospin or Nucleobond PC100 kits (Macherey-Nagel), depending on amount needed. Isolation of plasmid DNA for transfection of eukaryotic cells (for virus production or transient transfection) was always performed with the Nucleobond PC100 EF kit

RNA Isolation

All RNA-preparations for qRT-PCR were performed with the RNA-easy Mini Kit (Qiagen) according to manufacturer's instructions. Usually, four chMM cultures were washed with PBS and subsequently lysed in 300 to 550µl RLT buffer. RNA that was not purified immediately was frozen in liquid nitrogen after cell lysis in RLT buffer and stored at -80°C until further processing.
For RNA-seq, total RNA from chMM cultures was isolated using peqGOLD TriFast Reagent (peqLAB, Erlangen) and subsequently purified with the RNeasy Mini Kit (Qiagen). Extraction of mRNA from total RNA was performed using the Oligotex mRNA Mini Kit (Qiagen) and 6 µg of total RNA.

cDNA Synthesis

cDNA synthesis for qRT-PCR was performed with the TaqMan Reverse Transcription Reagents Kit (Applied Biosystems) using 1µg of total RNA as template and random hexamer primers. The reaction was performed in 100µl volume.

Polymerase Chain Reaction (PCR)

PCR was generally performed using in-house produced *Taq* and *Pfu* polymerases (A.C. Stiege) or Phusion DNA-polymerase (NEB) when amplifying protein-coding expression constructs.

Quantitative Realtime-PCR (qRT-PCR)

qRT-PCR enables quantification of template DNA in samples using the SYBR Green I chemistry (Applied Biosystems) and was run on an ABIPrism 7900 HT thermocycler. Primers were designed by Primer3Plus (Untergasser et al. 2007) or by hand and double-checked with the program OligoCalc (Kibbe 2007).

Relative quantification of transcript levels

To compare RNA abundance between various samples, cDNA was generated as described above. The qRT-PCR was set up in 18µl reaction volume (3µl Primer Mix (4.5pmol per primer); 6µl cDNA template; 9µl SYBR Green PCR Master Mix) on a 384-well plate. Each reaction was performed as a triplicate.

The standard curve was generated from a cDNA-pool of all samples that were measured in an experiment and was diluted in 1:2 steps (1 – 0.5 – 0.25 – 0. 125 – 0.0625). Relative values were measured for each unknown reaction and then normalized to the housekeeping gene GAPDH in each experiment, allowing relative quantification of changes in transcript abundance.

Absolute quantification

While relative quantification measures transcript levels relative to a housekeeping gene (GAPDH), absolute quantification aims at measuring the number of molecules in a sample. For this, "standard plasmids" carrying the amplicons of interest were cloned (see below). This allows calculation of the number of molecules given a known concentration of plasmid DNA. The standard curve was generated with the "standard plasmids" starting from a known concentration of 1 ng/µl and diluted 1:10 steps. The 10^{-2} to 10^{-6} dilutions were used as creation standards. In the highest concentration (10^{-2}) a PCR contained 0.06 ng plasmid DNA, the lowest concentration had 6 attograms plasmid DNA per PCR. To avoid loss of DNA at these low concentrations, all reactions were performed in nuclease-free H_2O to which 0.01% Tween20 was added. This greatly reduced initial inconsistency in PCR amplification.

Sanger Sequencing

The sequencing-PCR was performed with the BigDye v3.1 kit (Applied Biosystems), and 100ng plasmid-DNA as template. The product was cleaned by ethanol precipitation before transfer to the Charité sequencing facility for capillary electrophoresis, which was performed by Mohsen Karbasiyan on an ABI3700 capillary sequencer.

2.2.2 Cloning

RCASBP Expression Constructs

All wildtype and mutant *Hoxd13* RCASBP vectors used in this study were cloned by Asita Stiege in collaboration with Dr. Jochen Hecht. The constructs are based on the murine *Hoxd13* gene (NM008275.2 → NP032301.2) and carry a 3xFLAG tag at the N-terminal end of the protein. The RCASBP(A)-empty vectors were also available in the group.

In the course of the project four new RCASBP-vectors were created, carrying the chicken and murine version of the *PITX1/Pitx1* gene (NM_001167684.1 and NM_011097.2 respectively), preceded by a 3xFLAG-tag. Constructs for the RCASBP(A) and RCASBP(B) vectors were constructed in parallel, following identical cloning strategies, only differing by vector backbone (RCASBP(A) and RCASBP(B) respectively) in the final subcloning step.

The *cPITX1* CDS was amplified from chicken limb bud (HH25) cDNA with primers Gg_Pitx1_NcoI_F and Gg_Pitx1_SpeI_R and subsequently *Nco*I digested. The pSLAX13-5'3xFLAG vector was *Eco*RI digested, and blunted by T4-Polymerase, followed by an *Nco*I digestion step. This enabled a directional ligation of the *cPITX1* insert into the vector and creating a 3xFLAG tag using the *Nco*I site. This pSLAX13-3xFLAGcPITX1 shuttle vector was then digested with *Spe*I/*Cla*I to cut out the *cPITX1* construct, that was then ligated into the also *Spe*I/*Cla*I cut RCASBP(A) and (B) vectors respectively.

The *mPitx1* CDS was amplified from murine brain (E11.5) cDNA with primers Mm_Pitx1_NcoI_F and Mm_Pitx1_RI_F and was cloned into RCASBP(A) and RCASBP(B) following an almost identical cloning strategy. The only difference was that for the initial cloning step into pSLAX13-5'3XFLAG. Here PCR-product and vector were *Spe*I/*Nco*I digested before ligation.

Cloning of Standard Plasmids for Absolute Quantification

To generate standard plasmids for absolute quantification a PCR using the adequate primers was run on appropriate cDNA with *Taq* DNA polymerase. The PCR product was then purified after size selection through an agarose gel and ligated into a pTA-GFP vector using standard procedures.

Cloning of pGL3 Reporter Vectors

For luciferase assays, the *cSOX9* promoter sequence was cloned into the pGL3-basic vector (Invitrogen) using the *Xho*I/*Nhe*I restriction sites. The genomic sequences were amplified with primers (

Table 2.7) that had according restriction sites attached.

2.2.3 Biochemical Methods

Determination of Protein Concentration

Measurements of protein concentrations were performed using the BCA Protein Assay Reagent Kit (Pierce) according to manufacturer instructions. The absorption at 592nm wavelength was measured on a ELISA plate reader and quantified relative to a dilution series set up with BSA

SDS-PAGE

Gel electrophoresis for proteins (SDS-PAGE) was performed on 12.5% polyacrylamide gels as described in Sambrook et al 2001.

Western Blot

Protein Transfer

Western Blots were performed following standard procedures using a tank transfer system onto a PVDF membrane (Millipore Immobilon P, 0.45µm pore size). Transfer conditions included pre-cooled transfer buffer (25mM Tris-Base; 200mM Glycine, 20% Methanol) and a run time of 60-70 minutes at 100V in an ice-filled water bath or at 4°C.

Protein Detection

Immunodetection of antigens was performed as follows; 30 minutes blocking of the membrane in 5% milk powder in TBS-T (0.1% Tween-20); primary antibody incubation 1h at room temperature or 16h at 4°C; three TBS-T wash steps, 10 minutes each; secondary antibody incubation 1h at room temperature or 16h at 4°C; 3-5 wash TBS-T steps; Signal was detected using Western Lightning Plus ECL (Perkin-Elmer) and various X-Ray films, developed on the AGFA Curix 60.

Electromobility-Shift Assay (EMSA)

All EMSAs were performed by Asita C. Stiege and Dr. Jochen Hecht, who also produced and purified the recombinant HOXD13 homeodomains.

2.2.4 Cell Culture Methods

Transient Transfection and Luciferase Assays

NIH3T3 cells were grown in medium (DMEM 4.5% Glucose, 1% Penicillin, 1% L-Glutamine, 10% Fetal Bovine Serum) and seeded in 24-well plates with 2 x 10^4 cells per well. The following day cells were transfected using Polyfect transfection reagent (Qiagen) with 100 ng of pGL3-reporter vector and 200 ng of pTL10 expression vector. After 24 h of incubation cells were lysed with 5x Passive Lysis Buffer (Promega) and reporter activity was measured with DualGlo luciferase reagent (Promega) on a Trilux MicroBeta1450 scintillation counter. Each experiment was performed in triplicate and in three biological replicates.

Virus Production

Production of highly concentrated virus followed a protocol described in detail by Seemann (2006) . A modification to the described protocol was the initial used of Turbofect (Thermo Scientific) to transfect the DF1-cells with RCAS-plasmids instead of Lipofectamine.

Chicken Micromass Culture

The protocol used for preparation of chicken micromass (chMM) cultures has been adapted from (Lise et al. 2000) and modified by Petra Seemann (2006).

Eggs were incubated at 37.5°C and >60% humidity for 4.5days. Limb buds of fore- and hindlimbs from stage HH24-25 were dissected collected in room temperature PBS. Following 3-5 washes in prewarmed HBSS (Hanks' Balanced Salt Solution, Cambrex, Cat.-No.: BE10-547F) the limb buds were digested for 15min at 37°C in Dispase solution (Gibco, CatNo:17105-041, 3 mg/ml in HBSS) to detach the ectodermal from the mesenchymal layer and subsequently washed in prewarmed HBSS until the detached ectoderm was largely removed. The mesenchymal limb buds were then incubated in 1ml prewarmed digestion solution (0.1% (w/v) Collagenase type Ia (Sigma-Aldrich, Cat.-No. C9891), 0,1% (w/v) trypsin (Gibco), 5% FBS in PBS) for 30min at 37°C and gently pipetted up and down to create a single cell suspension. To these cells, 9ml of prewarmed chMM medium was added the suspension was filtered through a cell strainer (45µm, BD Bioscience) to ensure a single cell suspension. The cell strainer was washed with another 10ml of prewarmed chMM-medium.

The cells were counted (using a Neubauer cell counting chamber), pelleted and adjusted to a 2x10^7 cells/ml cell suspension (in chMM medium). Of this suspension a droplet of 10µl was seeded per well in a 24-well cell culture plate and incubated for 2h at 37°C to let the cells attach. Finally, 1ml of chMM medium was carefully added to each culture and medium was renewed every 2 days.

To exogenously express the TF of choice, chMM cultures were infected with the appropriate RCASBP(A) virus at the time of seeding. Approximately 0.5µl of concentrated virus was added per 10µl of $2x10^7$ cells/ml suspension before seeding (i.e. for a 24-well plate 12µl virus/240µl cell suspension). Of this virus/cell mixture 10µl droplets were seeded.

To generally assess the differentiation of the chMM culture, the individual cultures were stained with Eosin or Alcian Blue.

For eosin staining, cultures were fixed in 4%PFA in PBS at 4°C over night and subsequently washed in PBS. The cultures were stained with eosin solution for 40-90 seconds and after that briefly washed in PBS. Photographs were taken immediately afterwards without liquid inside the well, to prevent destaining.

For Alcian Blue staining, cultures were fixed for 1h room temperature or overnight at 4°C in Kahles Fixativ (30%EtOH, 0.4%PFA, 4% acetic acid) and subsequently incubated in After Eight Blue stain (0.05% Alcian Blue in 0.1M HCl-solution) over night at room temperature. The chMM cultures were washed with PBS or freshly drafted tab water to remove excess Alcian Blue stain and, as for eosin staining, photographed without liquid inside the well.

2.3 Chromatin-Immunoprecipitation (ChIP)

Chromatin Immunoprecipitation comprises a multistep protocol that enables detection of protein-DNA interactions. In brief, cells or tissues are fixed with formaldehyde to crosslink the protein DNA-complexes of the chromatin. The nuclear chromatin is then purified and fractioned by sonication. The sonicated chromatin is subjected to immunoprecipitation with antibodies against transcription factors, modified histones or other nuclear antigens. Subsequently, the co-precipitated DNA can be analyzed in various ways. Analysis by qRT-PCR for example, allows testing the enrichment of specific DNA fragments, compared to control regions. In this work, ChIP-purified DNA was used for massively parallel sequencing.

2.3.1 Cell Fixation

To harvest chMM cultures for ChIP, medium was aspirated and 200-400µl of collagenase-solution (0.1% (w/v) Collagenase type Ia (Sigma-Aldrich, Cat.-No. C9891) in chMM-medium) added. chMM cultures were incubated in collagenase-solution for 30 to 90 minutes and then disrupted by pipetting up and down. This creates smaller structures and allows a more homogenous fixation. Usually cells from 48 chMM cultures (two 24-well plates) were pooled for one biological replicate. All cultures were collected in a 50ml Falcon, pelleted and resuspended in 10ml cold chMM-medium.
To crosslink, 273.5µl 37% formaldehyde (Roth) was added to the cell suspension (1% final concentration), which was briefly inverted and incubated for 10 minutes on ice. After 10 minutes 555µl of 2.5M glycine (0.125M final concentration) was added to quench the formaldehyde fixation. Subsequently, cells were washed twice in cold PBS and, if not directly subjected to cell lysis, shock frozen in liquid nitrogen.

2.3.2 Nuclear Extraction

This protocol is a modified version of the ChIP-protocol published by Lee et al (2006). Cross-linked chMM cells were thawed (if appropriate), resuspended in 10ml cold Lysis Buffer 1 and incubated with gentle rocking for 10 minutes at 4°C. Cells were spun down (5min at 2700xg), the supernatant discarded, resuspended in 10ml Lysis Buffer 2, and incubated with gentle rocking for 10 minutes at room temperature. Cells/nuclei were spun down (5min at 2700xg), the supernatant discarded and resuspended in 1.5ml Lysis Buffer 3.

2.3.3 Sonication

The samples were sonicated with a Bioruptor NextGen for 35 to 45 cycles (40 cycles for most experiments) 30 seconds pulse, 30 seconds pause). During sonication, the tubes were immersed in a cooled water bath (4°C).

To remove cell debris, 150µl 10% TritonX100 ($1/10^{th}$ of the volume) were added to the samples and they were centrifuged for 15 min at 16,000xg and 4°C. The supernatant contained the sonicated chromatin, which was transferred to a fresh reaction tube.

2.3.4 Quality Control of Sonicated Chromatin

Before proceeding to ChIP, the quality control of each chromatin sample was performed. For this, 50 to 100µl of sonicated chromatin were transferred to a fresh reaction tube and mixed with $1/10^{th}$ of the volume 5M NaCl. Reversal of formaldehyde crosslinks took place by overnight incubation at 65°C. Subsequently, RNA was digested by adding 2µl RNase and incubating at 37°C for 30 minutes, followed by a proteinase K digest (2µl and 30 minute digestion at 55°C). The DNA was purified by ethanol precipitation and dissolved in 20µl nuclease free H_2O. Then, the concentration in the sample was measured and used to infer the chromatin concentration in the original sample.

$$\frac{\left[(concentration)\times(resuspensionVol.)\times\frac{(sonifiedChromatinVol.)}{(Vol.forQC)}\right]}{(sonifiedChromatinVolume)} = \mu g \text{ Chromatin in 1.6ml original sample}$$

Efficiency of sonication was checked on a 1% agarose gel where, in case of successful fragmentation, the majority of the DNA was expected to be sheared to fragments of 200-500bp length.

2.3.5 Immunoprecipitation

Prior to immunoprecipitation, protein G beads were blocked by washing two to three times with 1ml of Blocking Solution (0.5% BSA in PBS), and finally resuspended in 100 to 400µl Blocking Solution. Then, antibody was added to the protein G beads and incubated with gentle rocking at 4°C for at least 3 hours. After this the antibody-bead complexes were washed twice with Blocking Solution. 10µg of anti-FLAG antibody and 40µl of protein G beads were used for a standard ChIP experiment.

If the sample of sonicated chromatin was of sufficient quality, a volume equivalent to 35µg of chromatin was used as sample for ChIP. In parallel, 100 µl of sonicated chromatin were taken as input chromatin sample and stored at -20°C until crosslink reversal (see below).

To each 40µl of antibody-coated beads, a volume equivalent to 35µg of sonicated chromatin was added and incubated over night at 4°C with constant gentle rocking.

Afterwards beads were captured using a magnet and the supernatant was discarded and beads were washed six times with RIPA Buffer followed by one washing step with TE-NaCl buffer. All supernatants were discarded. Following the final wash step, the beads were centrifuged for three minutes at 1000 rpm and 4°C and the remaining few µl of supernatant were carefully aspirated. To elute the protein-DNA complexes from the beads, 210µl of Elution Buffer was added to the beads and incubated for 30 minutes at 65°C and 1200rpm on a thermomixer. Afterwards the beads were spun down (1 minute at 16,000 rpm), and 200µl supernatant, containing the immunoprecipitated protein-DNA complexes were transferred to a fresh reaction tube. At this point the 100µl input chromatin samples were thawed and processed in parallel to the ChIP samples.
Crosslink-reversal was performed by adding $1/10^{th}$ of the volume (20µl for ChIP, 10µl for input) 5M NaCl, and overnight incubation at 65°C. Following RNase A (4µl, 30min, 37°C) and proteinase K (4µl, 30min, 45°C) digestions, the DNA was purified by ethanol precipitation using 4µl of Glycogen (Ambion 5mg/ml, Cat.-No. AM9510) as a carrier.

The precipitated sample was sent to the BCRT sequencing facility for further processing. ChIP and input DNA samples were sequenced on an Illumina GAIIX sequencer in 36bp single-end reads.

2.3.6 Initial Processing of ChIP-seq Data

As mentioned earlier, ChIP–seq allows genome-wide detection of TF binding sites. Through massively parallel sequencing millions of short sequences, called reads, are generated from which one needs to extract the binding information. This requires multiple processing steps of the initial reads.
Briefly, the initially sequenced reads are quality filtered and subsequently aligned (mapped) to a template genome. Due to sonication, co-precipitation of protein-DNA complexes enriches for short (200 – 500bp) DNA-fragments encompassing the location where the protein has bound. Since only the 5' ends of these fragments are sequenced from one of either end, one expects an enrichment of reads in the genome surrounding a TF binding site. This enrichment can be detected after mapping the reads to the genome. It usually follows a very characteristic bimodal distribution of reads surrounding a binding site, where an enrichment of reads on the positive strand is followed by a similar enrichment on the negative strand (see Figure 1.5). When displaying the coverage of reads along the genome in a graph, the shape can resemble a mountain or mountain range, which is why an individual binding site has been termed peak. After mapping the reads to the genome, computer programs, aptly termed peak-callers, are used to automatically detect the characteristically enriched locations in the genome.

Each ChIP-seq experiment is performed in biological replicates (independently prepared chMM cultures and so forth). A crucial step is to validate the reproducibility of identified peaks. Only after validation of replicate reproducibility, a list of high-confidence binding sites is generated from the replicate experiments, which then can be further analyzed. The processing steps used in this study are based on the guidelines for ChIP-seq experiments published as part of the ENCODE publications (Landt et al. 2012).

All bioinformatics analysis was run on a server at the Charité and operated by using a locally installed version of the GALAXY platform (Blankenberg et al. 2010; Goecks et al. 2010). The infrastructure for bioinformatics processing at the Charité was set up and is maintained by Peter Hansen, who also installed all additional programs that were not part of the standard GALAXY tools.

Initial Quality Control and Read Mapping

Initial quality of sequences was assessed using the FastQC program and all reads discarded that had an average Phred-score < 28 using the Filter-by-Quality function from the FASTX-toolkit. The remaining reads were mapped to the chicken genome (WUGSC2.1/galGal3) with the *BWA aligner* and following parameters (aln -n=0, aln -o=1, aln -e=1, aln -d=16, aln -i=5, aln -l=-1, aln -k=2, aln -M=3, aln -O=11, aln -E=4, aln -R=FALSE, aln –N=FALSE, samse/sampe –n=3, sampe=–N10, sampe –a=500, sampe –o=100000, samse/sampe –r=NO). This created a .sam file of aligned reads. Next, all reads that aligned to more than one location were removed by *selecting lines* in the .sam file *matching* XT:A:U, @SQ, or @PG (XT:A:U is a tag in the .sam file that indicates a uniquely mapped read. @SG and @PG are comment lines). Next, the *rmdup* function from the *SAMtools* package was used to remove PCR artifacts from library preparation after the .sam file had been converted to a .bam format. Finally, the enrichment of the ChIP was assessed using the *cross-correlation analysis* of the *spp 1.11* package (get-binding characteristics function, parameters: -srange=0,1000; -bin=5; -cluster=2; -debug=F; -MinTagCount=1000; -Acceptance_Z_Score=3; -RemoveTagAnomalies=T; -Anomalies_Z=5; -AcceptAllTags=F).

After performing this initial processing, the overall quality of the enrichment through ChIP was evaluated. The ChIP-seq guidelines suggested a number of quality thresholds, although not all thresholds need to be met by each experiment. The most meaningful metrics are the RSC and NSC values that can be calculated from the cross-correlation analysis and give an approximation of the enrichment from the immunoprecipitation step.

Reproducibility Testing

For each ChIP-seq experiment two biological replicates were created. Initial quality control was performed independently for each experiment. Next, reproducibility of the genome-wide enrichment needed to be tested. For this, a protocol taking advantage of a statistical metric called *Irreproducible Discovery Rate* (IDR) (Li et al. 2011) was followed, which was developed in the ENCODE consortium to compare ChIP-seq replicates.

The IDR aims at comparing the similarity of two ChIP-seq peak lists with making as little assumptions as possible. It compares two ranked lists of peaks (strongest to weakest) making only one assumption: The higher the significance of a peak in one replicate, the more likely it is to be a highly significant peak in the other replicate. If the peak is a false positive, it is less likely to be detected in the other replicate. The IDR algorithm uses signal peaks (true positives) and noise peaks (false positives) to distinguish between the two categories. Subsequently, the IDR is used to determine a threshold for the number of peaks above a certain *irreproducible discovery rate* (IDR). These peaks are likely to represent a reproducible "true positive". Thereby, the method can be used to assess the reproducibility of two ChIP-seq experiments and to determine a threshold for significant peaks that is adjusted depending on data quality.

Table 2.14 Quality thresholds and metrics used for assessing ChIP-seq quality

Metric	Origin	Suggested threshold	Meaning
Sequencing depth	Number of uniquely mapped, non-redundant reads per experiment	>10 million reads	Sufficient number of reads to expect most binding sites recovered
NRF (non-redundancy fraction)	$\frac{\#\,non-redundant\ uniquely\ mapped\ reads}{\#\,uniquely\ mapped\ reads}$	≥ 0.8 for 10million reads	low NRF indicates that only little DNA was recovered after ChIP (low library complexity)
NSC (normalized strand coefficient)	$\frac{cc\ (fragment\ length)}{\min\ (cc)}$	≥ 1.05	Both metrics measure the signal-to-noise ratio of a given experiment.
RSC (relative strand correlation	$\frac{cc\ (fragment\ length) - \min\ (cc)}{cc\ (read)\ length) - \min\ (cc)}$	< 0.8	

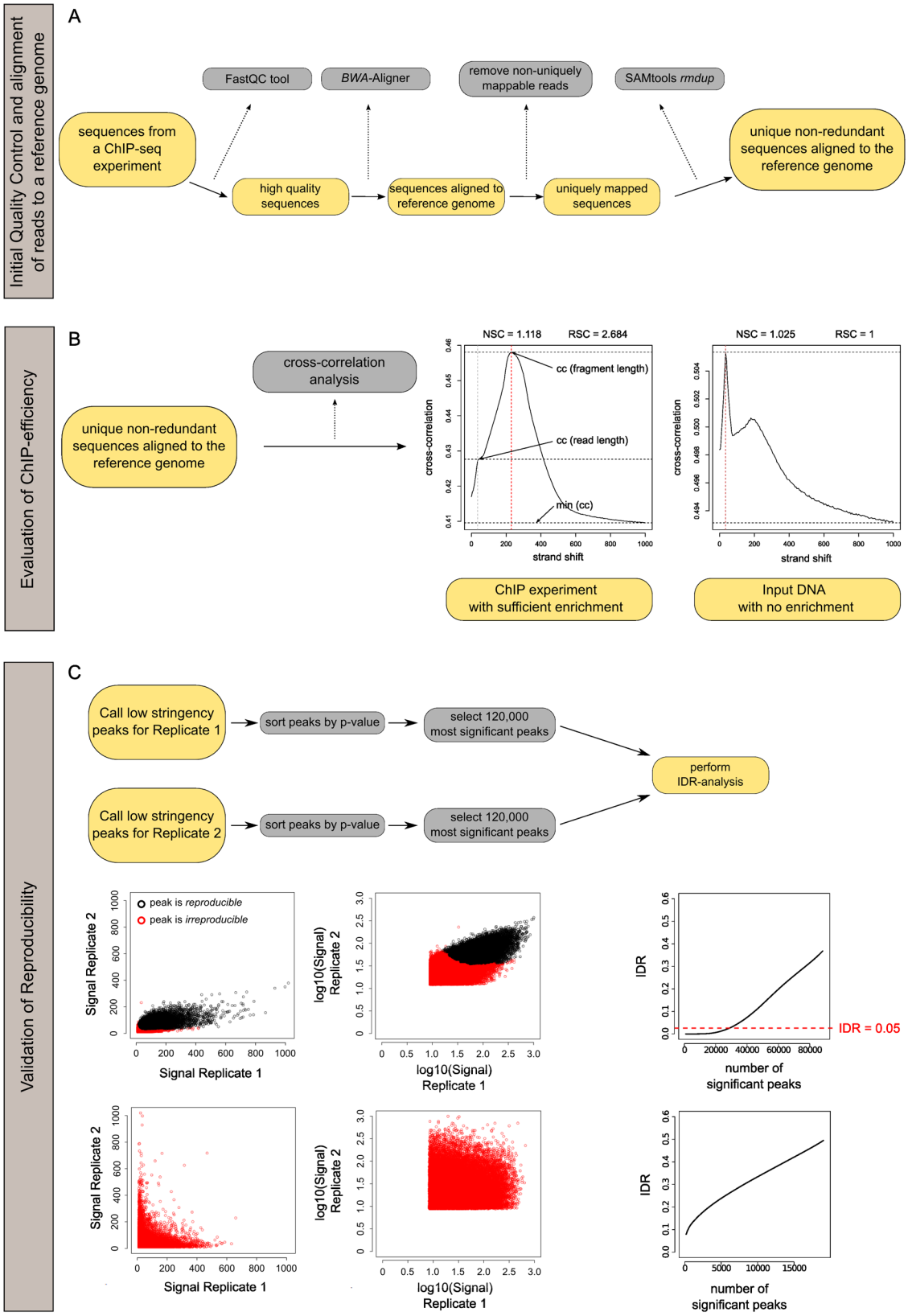


Figure 2.1 Quality Control steps and basic bioinformatics analysis of ChIP-seq data to validate enrichment and reproducibility of experiments. See text for Details

Peak Calling

All peaks were called against pooled input controls of both biological replicates with low stringency using the MACS2 program (parameters: -g 1.0e9, –p 1e-1, --too-large, --bw (as determined by spp in the quality control step)). The low-stringency peak calling generated lists of up to 500,000 peaks, consisting predominantly of false-positives. For subsequent IDR analysis, these lists were sorted by p-value and the 120,000 most significant peaks selected (even here, most peaks represent false-positives).

For IDR analysis, peak lists for each biological replicate were created, and additionally peak-lists for pseudo replicates[4] of each individual biological replicate and the pooled biological replicates. This comprises the eight peak-lists required for IDR analysis:

i) biological replicate 1 (Rep1)
ii) biological replicate 2 (Rep2)
iii) biological replicate 1 pseudorep. 1 (Rep1.1)
iv) biological replicate 1 pseudorep. 2 (Rep1.2)
v) biological replicate 2 pseudorep. 1 (Rep2.1)
vi) biological replicate 2 pseudorep. 2 (Rep2.2)
vii) pooled replicates pseudorep. 1 (Rep0.1)
viii) pooled replicates pseudorep. 2 (Rep0.2)

IDR-Analysis

To test reproducibility, IDR analysis is performed for the biological replicates (*replicate consistency*), the pseudo replicates of each biological replicate (*self-consistency*), and the pseudo replicates of the pooled replicates (*pooled-consistency*).

Each IDR analysis gives the number of peaks above a certain IDR threshold. This is, for a 0.01 (1%) threshold, the number of peaks that have a 99% probability to be reproducible between the replicates. For *replicate consistency* and *self-consistency* IDR thresholds of 0.01 were used, for *pooled- consistency* 0.0025.

Reproducibility of two datasets is ensured if the numbers of peaks above the threshold for the *self-consistency* sets are within the factor of two. Also, the number of peaks above threshold for *replicate-consistency* and *pooled_consistency* should be within a factor of two. More details on this method can be found in (Li et al. 2011; Kundaje 2012; Landt et al. 2012)

Determination of the Final Peak Set

The final peak-set was compiled by calling low-stringency peaks on the pooled reads from both replicates, which were then sorted by p-value. Of this list, the top number of peaks as given by the *pooled-consistency* analysis (peaks above IDR 0.0025) was selected and make up the high-confidence peaks for a given ChIP-seq experiment.

[4] Pseudo replicates are used to test the self-consistency of a dataset. Reads from a ChIP-seq experiment are randomly distributed onto two „pseudo replicates“ and peaks are called for each set independently. IDR analysis of the pseudo replicates is expected to be very good, since it effectively measures the same dataset twice. More importantly, it also shows how good the signal-to-noise ratio for an individual dataset is by comparing the reproducible peaks called with half of the input reads.

2.4 Bioinformatic Analyses

Motif analysis

DREME

Motif analysis was conducted using *DREME* (MEME package, version 4.7.0) (Bailey 2011). Sequences of 150bp surrounding the summits of the 1000 most significant peaks (sorted by p-value) were used as input (non-default parameters: -mink 3, -maxk 10).

RSAT

Motif analysis with RSAT was performed using the identical input sequences as for DREME. (non-default parameters: -disco oligos, positions -nmotifs 5 -minol 6 -maxol 7)

Peak overlap

To calculate peaks the peaks were set to peak summits +/-150bp, creating a list of uniform peaks with 300bp width. Two peaks were considered overlapping when the overlap would be at least 200bp in length. This meant, that the peak summits were not further than 100bp apart and that if three peaks overlap, each pair of peaks would also overlap.

seqMINER Analysis

The seqMINER release 1.3.3 (April 2012) (Ye et al. 2011) was used. The peak summits were taken as reference coordinates. Subsequently, mean read densities relative to these locations were calculated for pooled reads from the two biological replicates for each factor. The graph displays is the distribution of mean read densities (tags/50bp) centered on the peaks summits (+/- 3000bp).

Motif Tools

For further analysis of motif distribution in ChIP-seq peaks the FIMO and CENTRIMO algorithms of the MEME suite (Bailey et al. 2009) were used, unless otherwise stated with default parameters. The motif matrices used as input correspond to the motifs shown in the according figures. Initial motifs detected by RSAT or DREME were manually shortened, so that only the core motifs with high information content bases were represented.

RNA-seq

Dr. Marten Jäger and Peter Hansen at the Computational Biology Group of Prof. Dr. Peter Robinson at the Charité performed most processing and data analysis steps of RNA-seq data.

Primary RNA-seq processing steps (Dr. Marten Jäger)

The reference transcriptome for *G. gallus* was downloaded from Ensembl release 64 database (WASHUC2/GALGAL3 genome assembly) and contains 23,392 transcripts (exon models) for 17,934 genes. Reads for each dataset were separately mapped to the reference transcriptome using bowtie (release 0.12.7) mostly using default parameters (exceptions: seed length (*l=40*), the number of allowed mismatches in the seed region (*m=3*)). A preceding quality filtering of the reads was skipped due to bowtie's ability to use Phred quality scores of reads during the alignment step. Reads mapping to multiple transcripts were randomly assigned to one of them. To quantify the transcript expression and normalize between the datasets the *reads per kilobase of exon model per million mapped reads* (RPKM) values for all exon models were calculated (Mortazavi et al. 2008) and used for further analysis.

Hierarchical Clustering and Principle Component Analysis of Expression Data (Dr. Marten Jäger and Peter Hansen)

The analysis of expression data was performed for the five RNA-seq datasets with RPKM values for 3118 transcripts showing a minimum RPKM value of 2 in at least one of the experiments with the additional criterion of a fold change of at least two compared to the RPKM values obtained for the control RCASBP vector. Principle Component Analysis (PCA) and hierarchical clustering were applied to the base-2 logarithms of the RPKM fold change values. The PCA was performed using the R function prcomp (retx=TRUE, center=TRUE, scale=TRUE). The hierarchical clustering and dendrogram calculation was done using the Matlab 2011 function clustergram with the Euclidean metric for pairwise comparison and unweighted average distance for the linkage.

Gene Ontology Analysis (Daniel Ibrahim)

GO-analysis was performed using DAVID (Huang et al. 2008; Huang et al. 2009). Input lists were ENSGALG numbers found after expression filtering.

Co-Binding of TF to Differentially Co-regulated Genes (Peter Hansen)

The 3118 differentially regulated transcripts the RPKM values were summed up for each gene and condition. RPKM fold changes relative to control cultures were calculated as for transcripts. Genes with a twofold up-regulation for HOXD13^{Q317K} and PITX1 or a twofold down-regulation

for HOXD13^{Q317K} and PITX1 were defined as HOXD13^{Q317K}/PITX1 co-regulated genes. HOXD13^{Q317R}/PITX1 co-regulated genes were defined accordingly.

Shared peaks for HOXD13^{Q317K} and PITX1 and accordingly HOXD13^{Q317R} and PITX1 were assigned to genes if they fall in the promoter (-5kb to +2kb of TSS), the gene body or gene flanking regions (-25kb to +25kb) of annotated ENSEMBL genes. Genes with at least one shared peak for HOYD13^{Q317K} and PITX1 were defined as HOYD13^{Q317K}-PITX1 cobound genes. Hoxd13^{Q317R}-PITX1 cobound genes were defined accordingly.

3 Results

3.1 Two Novel *HOXD13* Mutations Cause Distinct Human Phenotypes

3.1.1 Two *HOXD13* Missense Mutations, p.Q317K and p.R298Q, are Associated with Distinct Limb Malformations

Patients carrying two missense mutations in *HOXD13* causing distinct limb malformations were identified and initially described in the Institute for Medical and Human Genetics, Charité Berlin.

One case concerned a girl with a complex brachydactyly/oligodactyly phenotype. The girl was born in a family with no history of congenital malformations and with otherwise normal measurements and psychomotor development. On both hands the girl had only four, severely shortened fingers. The terminal phalanges as well as the nails of the thumbs were missing (Figure 3.1A, left). Radiograms of the right hand, performed at the age of five months showed only four metacarpals and only the proximal phalanges of the thumb and the two following digits were present (Figure 3.1A, right). In the little fingers, which also showed clinodactyly, the proximal phalanges and a hypoplastic middle phalanx could be identified. All other phalanges were missing (Figure 3.1).

The left and right foot had three and four toes, respectively, and the nails of both big toes were absent. As for the hand, the feet showed three and four metatarsals and only rudimentary proximal phalanges, all distal ones missing. Finally, the right foot displayed a syndactyly between the two central toes.

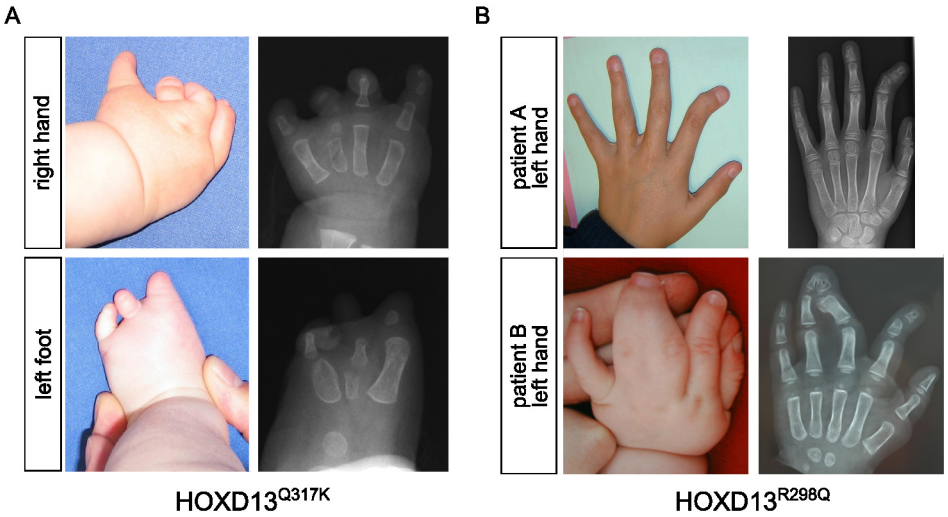


Figure 3.1 Two *HOXD13* missense mutations cause distinct phenotypes.

Photo- and radiographs of *HOXD13* mutant patients. (A) Hand and foot phenotype of the *HOXD13*^{Q317K} patient. Missing digits and absent distal phalanges and nails are apparent. Radiographs show missing and deformed metacarpals. (B) Left hands of the two HOXD13^{R298Q} patients. Note the BDA2 of the first digit with the notched terminal phalange. The SPD with an additional phalangeal anlage as seen in patient B was surgically removed in patient A.

The mutation underlying this severe phenotype was found to be a heterozygous c.949C>A mutation in exon 2 of the *HOXD13* gene (NM_000523.2). The mutation has not been reported in the dbSNP database or elsewhere and was not present in either of the parents and had thus occurred *de novo*. On protein level, the mutation converts the glutamine at position 317 to a lysine (p.Q317K) (Figure 3.2A). Notably, the affected residue corresponds to position 50 of the DNA binding homeodomain, which is crucial for binding site specificity (Figure 3.2B). Interestingly, a different point mutation (p. Q317R) has been reported previously at this position in a family with a very different phenotype consisting of a form of brachysyndactyly type V (Zhao et al. 2007).

The second mutation was also found as a *de novo* mutation in two non-related patients, both displaying a similar phenotype consisting of a combination of central polydactyly with 3-4 finger syndactyly (SPD) and brachydactyly type A2 (BDA2) (Figure 3.1B). The patients were bilaterally affected by the SPD and the BDA2. Patient A's SPD had been treated surgically and there is no information regarding a possible polydactyly. Radiographs of patient B's right hand showed an additional phalangeal anlage between the third and fourth finger, while the syndactyly in the left hand showed no polydactyly (Figure 3.2B). All patients displayed bilaterally affected index fingers with hypoplastic middle phalanges leading to a radial deviation of the terminal phalanges of the index fingers (Figure 3.1B). Moreover, the terminal phalanges of the index fingers display a very mild form of polydactyly, which can be seen as notched terminal phalanx on radiographs. The BDA2 phenotype was also noted for the 2nd toes of patient A while no syndactylies of the toes were reported. No foot malformations were reported for

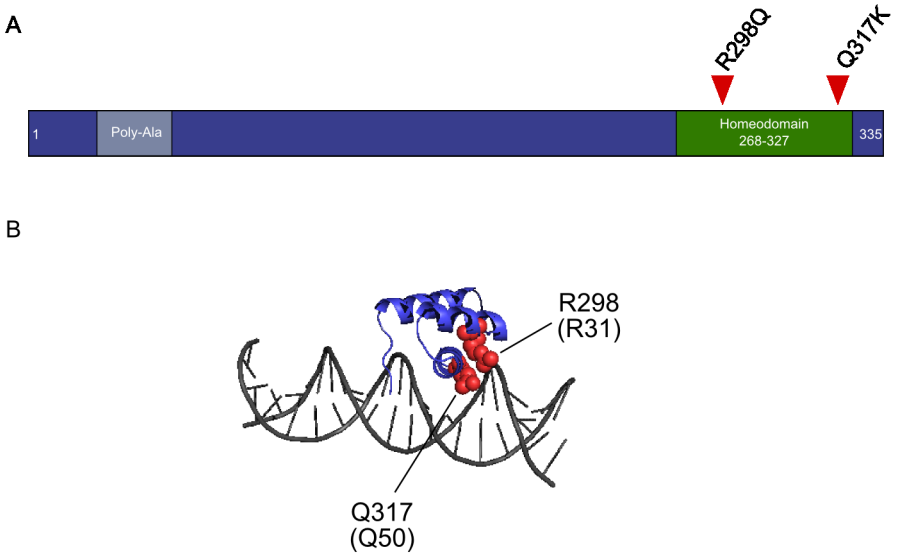


Figure 3.2 Location of the two HOXD13 missense mutations.

(A) Schematic representation of human HOXD13 indicating the known protein domains and position of the described missense mutations. (B) Three-dimensional structure of a homeodomain binding to DNA (*Drosophila* Engrailed (Fraenkel et al. 1998)). The arginine and glutamine mutated in the patients are highlighted in red and the atoms of the side chains are shown as spheres.

patient B.
Although novel at the beginning of this work, the same mutation causing an identical phenotype was reported by Wang and colleagues (2012a) in a two-generation Chinese family. In all three cases, the mutation responsible for the disease is a c.893G>A exchange in exon 2 of *HOXD13*. It converts the arginine at position 298 to a glutamine (p.R298Q) (Figure 3.2A). The affected residue, R298, is conserved in most homeodomains and has been described to mediate non-specific protein-DNA contexts through ionic bonds between the polar NH_2 groups of the side chain and the DNA phosphate backbone (Gehring et al. 1994; Berger et al. 2008) (Figure 3.2B). Another missense mutation also altering R298 was reported by Debeer et al. (2002). This R298W mutation causes a mild form of SPD with reduced penetrance. Affected patients, however, do not display the BDA2 phenotype seen in HOXD13^{R298Q} patients.

3.1.2 Functional Analysis Using Standard Biochemical Assays

First, I used standard biochemical and molecular biology assays to analyze the effects of the missense mutations on HOXD13 function.

I. Nuclear localization of HOXD13 mutants

To test whether the mutations affect nuclear localization of the HOXD13 proteins, as has been described for the poly-Alanine extensions in HOXD13, DF1 cells were infected with RCASBP(A) viruses expressing 3xFLAG-tagged versions of murine Hoxd13, that coded for wildtype, R298Q, and Q317K mutant proteins respectively. Figure 3.3A shows immunohistological staining of the cells with anti-FLAG antibody three days after infection. All variant HOXD13 proteins localize to the nucleus, indicating that the mutations do not affect the subcellular localization of the protein.

II. Binding of mutant homeodomains to the HOXD13 motif (EMSA)

Since both mutations affect the DNA-binding domain, the binding of the homeodomain to the known HOXD13 recognition sequence was tested. Wildtype and mutant HOXD13 homeodomains were recombinantly expressed in *E. coli* and subsequently purified. These were then used for electromobility shift assays (EMSA) using equal amounts (400 ng) of purified homeodomains and labeled DNA-oligos containing the HOXD13 recognition sequence published by Berger et al. (2008). Adding either unlabeled wildtype competitor oligos, which also carried the HOXD13 binding site, or mutant competitor oligos, which contained a modified HOXD13 binding site, tested sequence specificity of the DNA-binding.

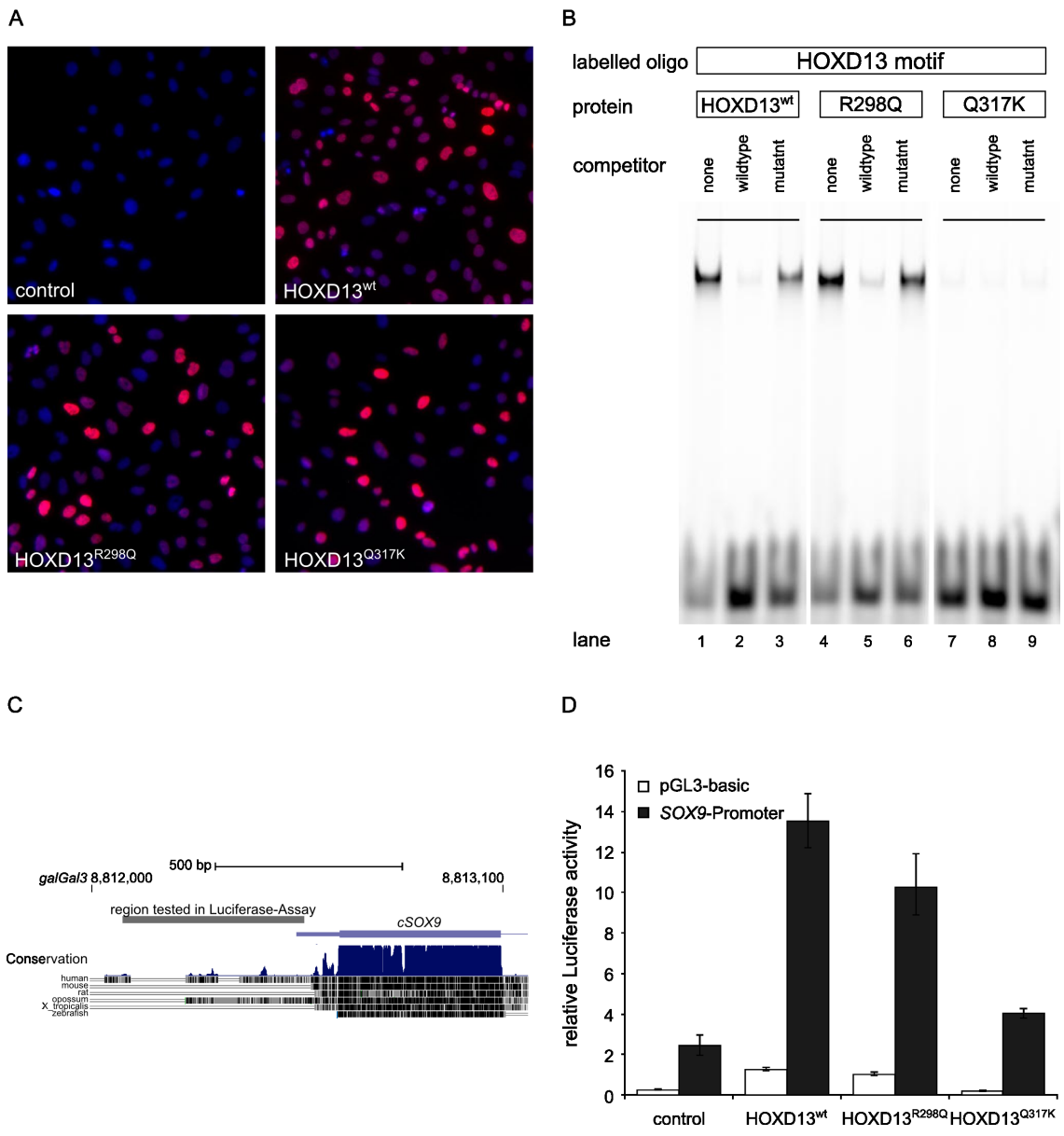


Figure 3.3 Standard functional analysis of HOXD13^{R298Q} and HOXD13^{Q317K} mutations.

(A) Immunocytochemical detection of virally infected DF1 cells shows the nuclear localization of wildtype and mutant HOXD13 proteins. DAPI (blue) stains the nuclei while an anti-FLAG antibody (red) detects the nuclear localization of the proteins. (B) *In vitro* binding properties of wildtype and mutant HOXD13 homeodomains. EMSAs were performed using fluorescent dye labeled double-stranded oligonucleotides containing a binding site (CCAATAAAA) for HOXD13. The oligos were incubated with purified wildtype and mutant homeodomains. Unlabeled competitor reduced binding to the labeled oligos whereas a mutant competitor did not, indicating sequence-specific binding. The HOXD13 recognition sequence is bound by HOXD13^{wt} and HOXD13^{R298Q} homeodomains but not by HOXD13^{Q317K}. (C) Schematic view of the promoter region used in luciferase reporter assays. (D) Transcriptional activation of the chicken *SOX9*-promoter by HOXD13 wildtype and mutants. NIH3T3 cells were transfected with 250 ng of reporter construct and 200 ng of *Hoxd13* cloned into expression vector pTL10. Bars represent the relative luciferase activity of at least three independent experiments. Co-transfection of *Hoxd13^{wt}* leads to approximately 5-fold activation of the reporter vector. *Hoxd13^{R298Q}* still achieves a 4-fold activation while *Hoxd13^{Q317K}* only shows a weak 1.6-fold stronger activation than with co-transfection of an empty pTL10 vector.

The wildtype HOXD13 homeodomain binds specifically to the labeled DNA-oligo carrying the HOXD13 binding site (band shift lane 1, Figure 3.3B). Adding excess unlabeled wildtype competitor removed the shifted band (lane 2). This was not the case when adding a mutant competitor oligo (lane 3), indicating sequence specificity of the protein-DNA interaction. The R298Q-mutant homeodomain showed an unchanged binding behavior and also specifically recognized the HOXD13 binding site (lane 4 – 6). In stark contrast, the HOXD13^{Q317K} mutation virtually abolished binding to the HOXD13 binding site (lane 7 – 9).

III. Activation of a luciferase reporter gene

Previous experiments in the group had indicated a regulation of the chondrocyte specific TF *SOX9* through *HOXD13*. To test the direct regulation of *SOX9* by HOXD13, the chicken *SOX9* promoter was cloned into a luciferase reporter vector (Figure 3.3C). Reporter assays carried out in NIH3T3 cells showed a 5-fold activation of the reporter construct when cotransfected with *Hoxd13*^{wt} (Figure 3.3D).

This *cSOX9* reporter vector was then used to compare the regulatory capacity of the HOXD13^{R298Q} and HOXD13^{Q317K} mutants to HOXD13^{wt} (Figure 3.3D). The HOXD13^{R298Q}-mutant activated the *cSOX9*-promoter although less effective than HOXD13^{wt} (approximately 20% reduction). The HOXD13^{Q317K} mutant fully lost its activating capacity and did not activate the reporter construct.

Taken together, standard approaches to analyze the effects of the missense mutations on HOXD13 function found the HOXD13^{R298Q} mutation to act as a hypomorphic allele with considerable amounts of residual wildtype activity, whereas the HOXD13^{Q317K} allele appears to act as a strong loss-of-function allele.

3.2 A Method to Functionally Characterize TF Mutations Using ChIP-seq

Using the standard approaches described above, varying loss of wildtype function could be observed, but – regarding genotype-phenotype correlation – led to presumably incomplete and inconsistent results. Therefore, ChIP-seq was chosen to allow a more comprehensive analysis of the HOXD13 mutations.

Since no specific antibodies that could differentiate between wildtype and mutant HOXD13 were available, chMM-ChIP-seq, a cell culture based system using mesenchymal limb bud cells and retroviral overexpression, was developed to overcome this hurdles. To be able to perform comparative ChIP-seq, murine sequences for *Hoxd13*^{wt}, *Hoxd13*^{Q317K}, and *Hoxd13*^{R298Q} were cloned into a retroviral vector, RCASBP(A), adding a 3xFLAG-tag to the N-terminus of the protein (Figure 3.4).

3.2.1 Validation of the Method

The system is based on chicken limb bud micromass cultures (chMM), which are generated from limb bud mesenchyme in which *HOXD13* is expressed physiologically. In culture, the cells spontaneously undergo a chondrogenic differentiation process that can be influenced by virally expressing a protein of interest using the avian-specific RCAS-system. Furthermore, previous studies in the group demonstrated that viral expression of *Hoxd13* inhibits chondrogenic differentiation in this context (Kuss et al. 2009).

To exclude that the 3xFLAG-tag affects HOXD13 protein function, the effect of tagged

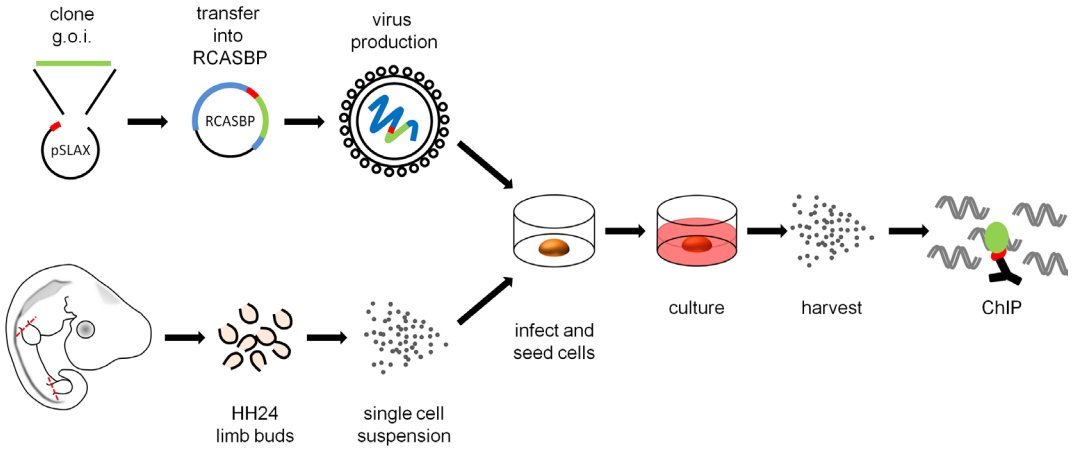


Figure 3.4 A system to stdy binding profiles of transcription factors

In vitro investigation o genome-wide TF binding using chicken mesenchymal cells and RCASBP retrovirus. The coding sequence of a gene of interest (g.o.i.) is introduced into the pSLAX vector in frame with a 3xFLAG tag sequence. This is transferred into an RCASBP virus using standard procedures. Single cell suspensions are isolated from the limb buds of chicken embryos at stage HH24 and are infected with the virus, leading to moderate overexpression of the protein of interest. After the desired culture time, cells are harvested and ChIP-seq is performed using an antibody against the FLAG-tag sequence.

HOXD13 was compared to an RCAS-virus expressing an untagged version (Albrecht et al. 2004). Chicken micromass cultures infected with the 3xFLAG-tagged version of HOXD13 had the same inhibitory effect on chondrogenic differentiation as untagged version (Figure 3.5A). For further molecular validation, expression analysis of selected *Hoxd13* target genes was performed comparing chMM cultures infected with untagged or 3xFLAG-tagged *Hoxd13* viruses. In line with the gross morphological effect, qRT-PCR analysis showed nearly identical up-regulation of *BMP4*, *FRAS1* and *MSX2* in both chMM cultures (Figure 3.5B).

Next, the degree of virally produced *Hoxd13*-transcript was compared to the endogenous transcript levels in posterior HH25 chicken limb buds, since viral overexpression can pose a problem by leading to excess amounts of TF protein,

All virally produced transcripts and all *Hoxd13*-coding transcripts were measured in chMM cultures using absolute quantification by qRT-PCR and set into relation to absolute *GAPDH* transcripts numbers, which were measured in parallel. Correspondingly, endogenous *HOXD13* transcripts were measured from HH25 posterior chicken limb buds and set into relation with *GAPDH* (Figure 3.6).

As shown in Figure 3.6, only 20% of all viral RNAs are spliced in such a way that a HOXD13 protein is translated, which is consistent with the published literature (Morgan and Fekete 1996)[5]. When comparing viral *Hoxd13* levels in chMM to HH25 posterior forelimb buds, chMM cultures showed moderate 3-fold higher expression of *Hoxd13* in chMM than *HOXD13* expression in limb buds (Figure 3.6).

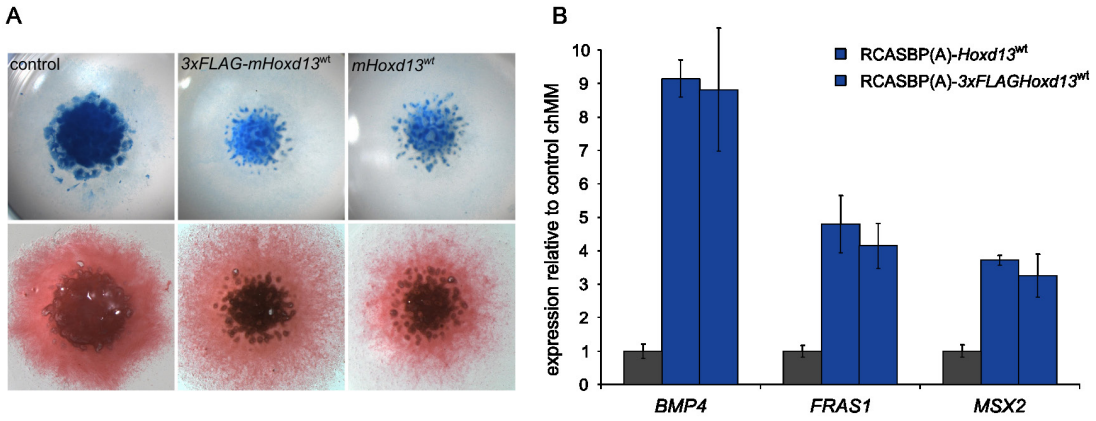


Figure 3.5 The 3xFLAG-tag does not affect HOXD13 function

(A) chMM cultures infected with virus carrying no insert, 3xFLAG-tagged *Hoxd13* or untagged *Hoxd13*. Inhibition of chondrogenic differentiation is indistinguishable between the two *Hoxd13* constructs. (B) qRT-PCR analysis of three genes, which are specifically up-regulated by Hoxd13. The bars indicated the relative expression of *BMP4*, *FRAS1* and *MSX2* in infected chMM cultures relative to expression in mock-infected cultures. Both *Hoxd13* constructs increase the gene expression to the same degree.

[5] This number was the same for $Hoxd13^{wt}$ virus and all tested *Hoxd13* mutant viruses. Also, $Hoxd13^{wt}$ and mutant infected chMM cultures produced comparable protein levels and showed no indication of protein degradation.

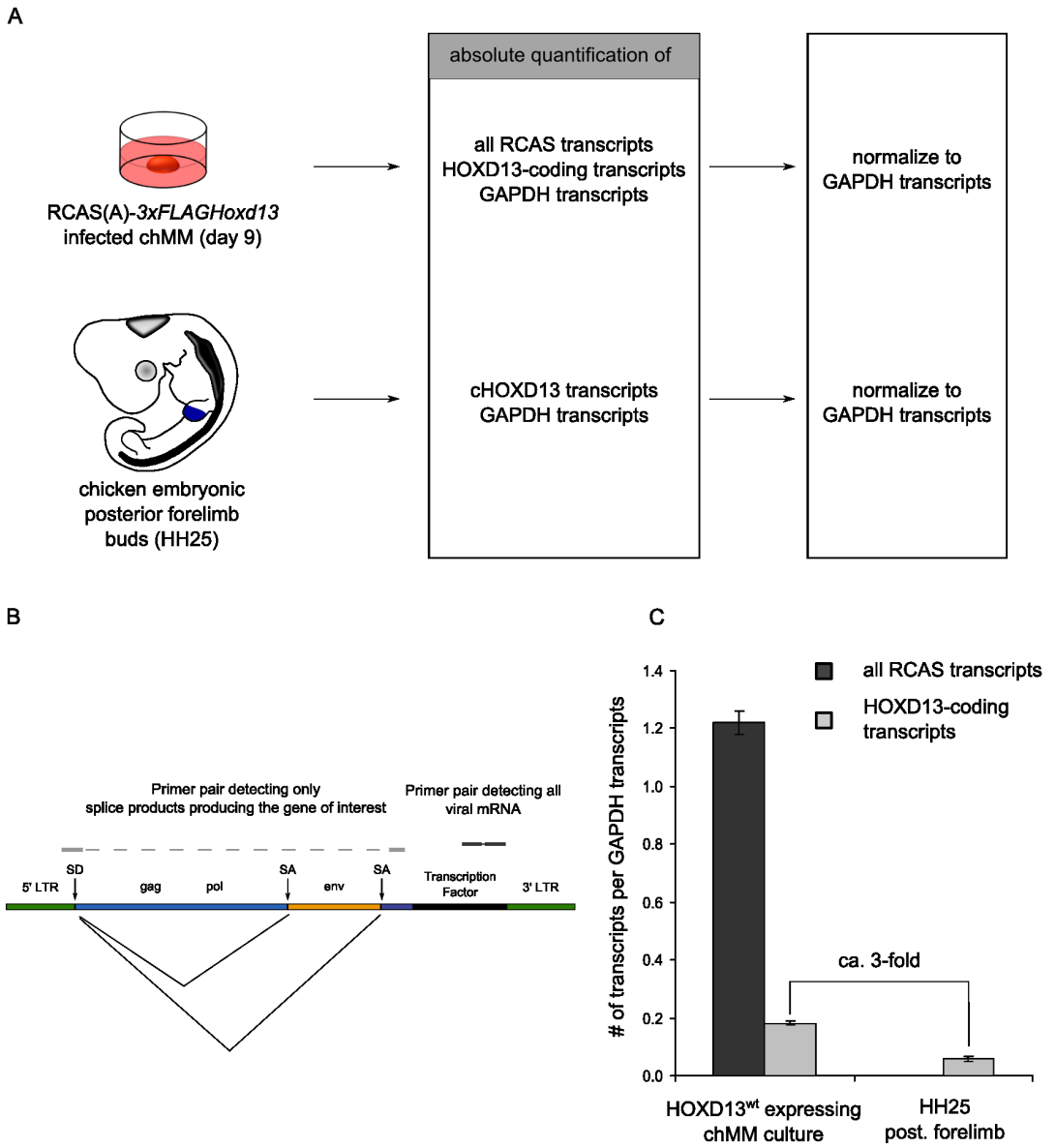


Figure 3.6 RCASBP-mediated expression induces only moderate overexpression

(A) Schematic overview of the approach to measure *Hoxd13* expression levels. Using absolute quantification of transcripts, the number of all RCAS-transcripts, *Hoxd13*-coding transcripts and *GAPDH* transcripts is measured in chMM cultures. For comparison, *HOXD13* and *GAPDH* transcript levels are measured in HH25 posterior chicken forelimb buds. Then, transcript levels are normalized to the number of GAPDH transcripts. (B) Schematic view of RCASBP viral RNA. Alternative splicing of the polycistronic RNA causes only a fraction of all viral RNA to code for the gene of interest. Position of the primer pairs that distinguish between gene of interest-specific and all viral RNA are indicated. (SD: splice donor site, SA: splice acceptor site, LTR: long terminal repeats) (C) Quantification of all viral RNA (black bar) and *Hoxd13*-coding RNA (gray bar) in chMM cultures. Only one in five viral transcripts codes for *Hoxd13*. Comparison to HH25 posterior forelimb buds shows that the RCASBP-induced overexpression accounts to approximately 3 to 5-fold overexpression. Error bars indicated the S.D. of three biological replicates (for chMM) and the S.D. of three technical replicates for HH25 forelimb buds.

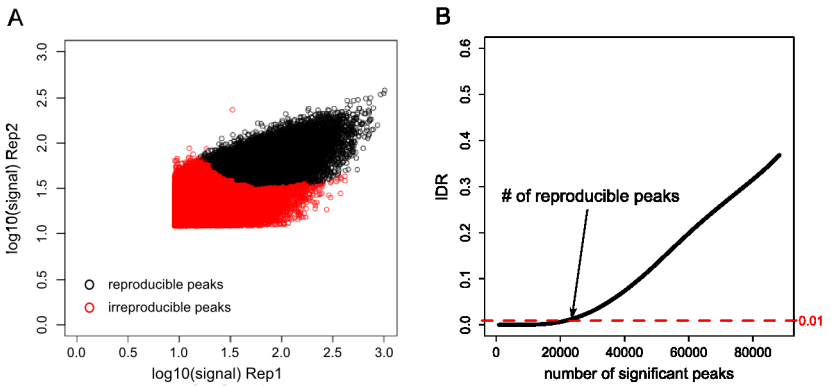


Figure 3.7 HOXD13^{wt} ChIP-seq reproducibility by the irreproducible discovery rate (IDR) framework.

(A) Scatterplots of signal scores for peaks that overlap in both replicates. Black circles represent peak pairs that pass an IDR threshold of 1%. Red circles represent peak pairs that do not pass the IDR threshold of 1%. (B) The estimated IDR as a function of different rank thresholds. The HOXD13^{wt} replicates show high reproducibility with 34,267 peaks passing an IDR threshold of 1%. See also Figure 2.1.

3.2.2 HOXD13-Binding in chMM

To study the genome-wide binding of HOXD13, chMM cultures infected with RCASBP(A) viruses expressing *3xFLAG-Hoxd13* were prepared in two independent replicates from which replicate ChIP-seq experiments were performed. Since the primary result of each ChIP-seq experiment are millions of short sequence reads, validation of reproducibility is no trivial task. To assess the reproducibility of ChIP-seq replicates, we[6] used the IDR method developed within the ENCODE consortium (Li et al. 2011; Landt et al. 2012), which specifically addresses reproducibility of ChIP-seq experiments (for details, see Section 2.3.6 and Figure 2.1). The two biological replicates for HOXD13^{wt} demonstrated good reproducibility in IDR-analysis, passing all criteria for consistent ChIP-seq experiments (Figure 3.7 and Section 5). In addition to evaluating the reproducibility of ChIP-seq experiments, the IDR method also suggests a threshold for the number of reproducibly detectable binding sites that can be identified from the given data. Using these criteria, 34,267 binding sites for HOXD13^{wt} were identified in chMM (Figure 3.7B).

To date, no ChIP-seq dataset for HOXD13 has been published. To validate the specificity of the identified binding sites, *de novo* motif analysis using sequences from the 1000 most significant peaks was performed with two different motif analysis tools (DREME and RSAT).

Both tools identified similar primary motifs for HOXD13^{wt} that contain a AT[A/T]AAA core sequence characteristic for HOX proteins, preceded 5' by a somewhat more loosely defined [C/T][A/C] (

[6] After several unsuccessful attempts to prove the reproducibility of our ChIP-seq experiments, Peter Hansen, in collaboration with whom much of the bioinformatic analyses were performed, suggested implementing the ENCODE data reproducibility criteria for our data. He was critical for progressing from successfully conducted wet-lab experiments to proper bioinformatic analysis of the ChIP-seq results.

Figure 3.8). Moreover, the ChIP-seq identified motifs were highly similar to the *in vitro* detected HOXD13 motifs (Berger et al. 2008; Jolma et al. 2013) and appeared more frequently in the Top 1000 peaks than the published motifs (Figure 3.8). This confirmed that using the chMM-based ChIP-seq approach allows reproducible detection of HOXD13 binding sites reflecting the proteins DNA-binding specificity and that it thus can be adapted to identify differences between wildtype and mutant HOXD13 proteins.

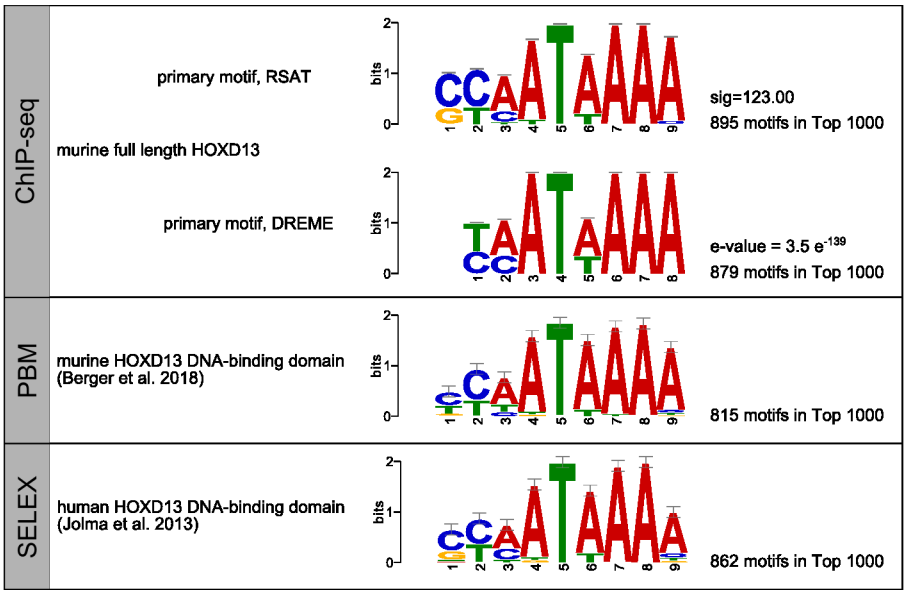


Figure 3.8 *De novo* motif analysis from ChIP-seq peaks identifies the canonical HOXD13 motif.

De novo motif analysis using sequences for the 1000 most significant HOXD13^{wt} peaks (summit +/- 75bp) as inferred by RSAT and DREME. Sig and e-values indicate the significance of the motif. The numbers refer to the number of motifs found in the Top 1000 peaks (FIMO p <0.0001). (FIMO p<0.0001). Comparison to motifs determined by Protein Binding Microarrays (PBM) (Berger et al. 2008) and high-throughput SELEX-seq (Jolma et al. 2013). Motifs were drawn using the MEME suite (Bailey et al. 2009).

3.3 Functional Characterization of the HOXD13^{Q317K} and HOXD13^{Q317R} Mutations

3.3.1 Binding Specificity

Initial experiments showed that the HOXD13^{Q317K} mutant localizes to the nucleus, but EMSA and reporter assays found the variant to act as a severe loss-of-function mutation. These results did not allow conclusions as to why the patient phenotype differs from other missense mutations in the homeodomain that also act as a loss-of-function in similar reporter assays.

To gain insight into the molecular pathomechanism underlying the *HOXD13^{Q317K}* mutation, the chMM-ChIP-seq approach was used to compare the binding properties of the HOXD13^{Q317K} mutant to HOXD13^{wt}, using the HOXD13^{Q317R} mutant as a control to further distinguish between specific amino acid changes at the Q317 residue.

Cross-linked chromatin from chMM cultures expressing tagged murine *HOXD13^{wt}*, *HOXD13^{Q317K}*, or *HOXD13^{Q317R}* were used for ChIP-seq and, after quality control and reproducibility testing, 34,267 binding sites for HOXD13^{wt}, 25,036 sites for HOXD13^{Q317K}, and 21,396 for HOXD13^{Q317R} were identified (see Section 5). To initially explore differences in binding specificity, *de novo* motif analysis was performed. Strikingly, primary motifs for both mutants revealed a changed 5' terminus of the motif (Figure 3.9). The primary motif for HOXD13^{Q317K} inferred by RSAT begins with a [G/A]G dinucleotide at the 5' end, followed by a variable base. The DREME motif is very similar but starts with single G a instead of a [G/A]G

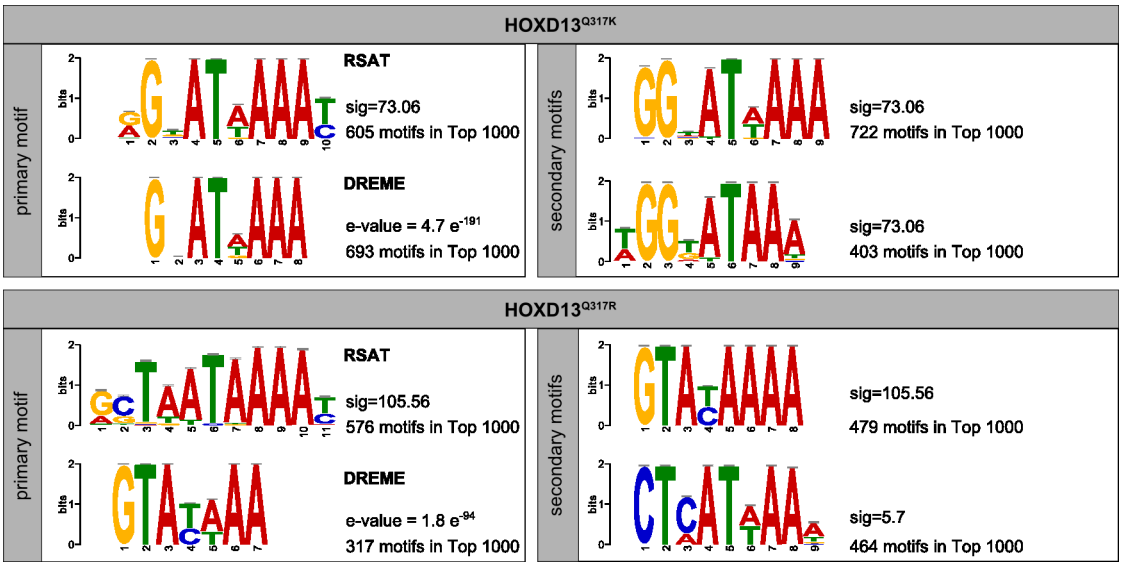


Figure 3.9 The HOXD13^{Q317} mutations alter the recognition site at the 5' terminus

De novo motif analysis using sequences for the 1000 most significant mutant HOXD13 peaks (summit +/- 75bp), as inferred by RSAT and DREME. The HOXD13^{Q317K} mutation recognizes a GGN trinucleotide at the 5' terminus. Selected secondary motifs (RSAT) show variations of the primary motif. The HOXD13^{Q317R} mutation recognizes a [G/A][C/G]TA at the 5' terminus. Selected secondary motifs (RSAT) for HOXD13^{Q317R} show a refined version of the primary motif or a CT[C/A]A 5' end, which is included in a subset of HOXD13^{wt} recognition sequences. Sig and e-values indicate the significance of the motif. Numbers are given for motifs found in the Top1000 peaks (FIMO p<0.0001). Motifs were drawn using the MEME suite (Bailey et al. 2009).

dinucleotide. The HOXD13^{Q317R} motif (RSAT) also starts with a [G/A][C/G], but is then followed by a thymine residue (Figure 3.9B). As for HOXD13^{Q317K}, the primary HOXD13^{Q317R} motif from DREME is similar but shortened at the 5' end, which starts with a single G followed by the thymidine residue.

Analysis of the secondary motifs identified by RSAT and DREME revealed further differences between HOXD13^{Q317K} and HOXD13^{Q317R}. The HOXD13^{Q317K} secondary motifs are comprised of a variety of motifs similar to the primary motif identified and strikingly feature a GG dinucleotide at the start. The HOXD13^{Q317R} secondary motifs, on the other hand, also resemble the primary motif with the characteristic GT dinucleotide in the beginning. In addition, both motif tools identified a secondary motif that is similar to the HOXD13^{wt} motif. This motif does not start with a GT dinucleotide but with a CT[C/A] followed by the AT[A/T]AAA core (Figure 3.9).

Taken together, primary and secondary motifs in both mutants clearly show an altered 5' terminus, whereas the AT[A/T]AAA core remains largely unchanged. This observation strongly suggests that Q317 mutations alter only the 5' terminus of the DNA recognition sequence of HOXD13 and do not affect the AT[A/T]AAA core of the recognition sequence.

The characteristic change in sequence specificity confirmed the reported role of the mutated residue (position 50 of the homeodomain, Q50) in sequence recognition. In order to identify other TFs with related recognition sequences, the HOXD13 homeodomain sequence was compared to all other human homeodomains in the Uniprot database carrying an arginine (R) or a lysine (K) at position 50. Eight R50 and 20 K50 homeodomain proteins are annotated in the human genome. The eight R50 matches are unusual for homeodomain-proteins in general, and none of the eight R50 homeodomain proteins is a developmental TF. Five R50-proteins are ceramide synthases (*CERS2-6*); enzymes bound to the endoplasmatic reticulum membrane (Levy and Futerman 2010). Another, *ARGFX* is a very weakly expressed transcript, and likely a pseudogene (Li and Holland 2010) and the two remaining, *ZFHX3* and *ZFHX4*, are tumor suppressor genes with multiple zinc-fingers and three full homeodomains, only one of harbors an R50 (Sun et al. 2005).

In contrast, the 20 K50 homeodomain proteins belong to various TF-families involved in different developmental processes (Figure 3.10A) and sequence alignment between these homeodomains and the mutant HOXD13 homeodomain revealed the greatest similarity to the PITX-family (Figure 3.10A). Furthermore, PITX1 is the only homeodomain protein carrying a K or R at position 50 known to be involved in limb patterning. PITX1 is a TF that determines hindlimb identity (Logan and Tabin 1999; Szeto et al. 1999) and has been shown to specifically bind to GGATTA (Tremblay et al. 1998)(Berger et al. 2008).

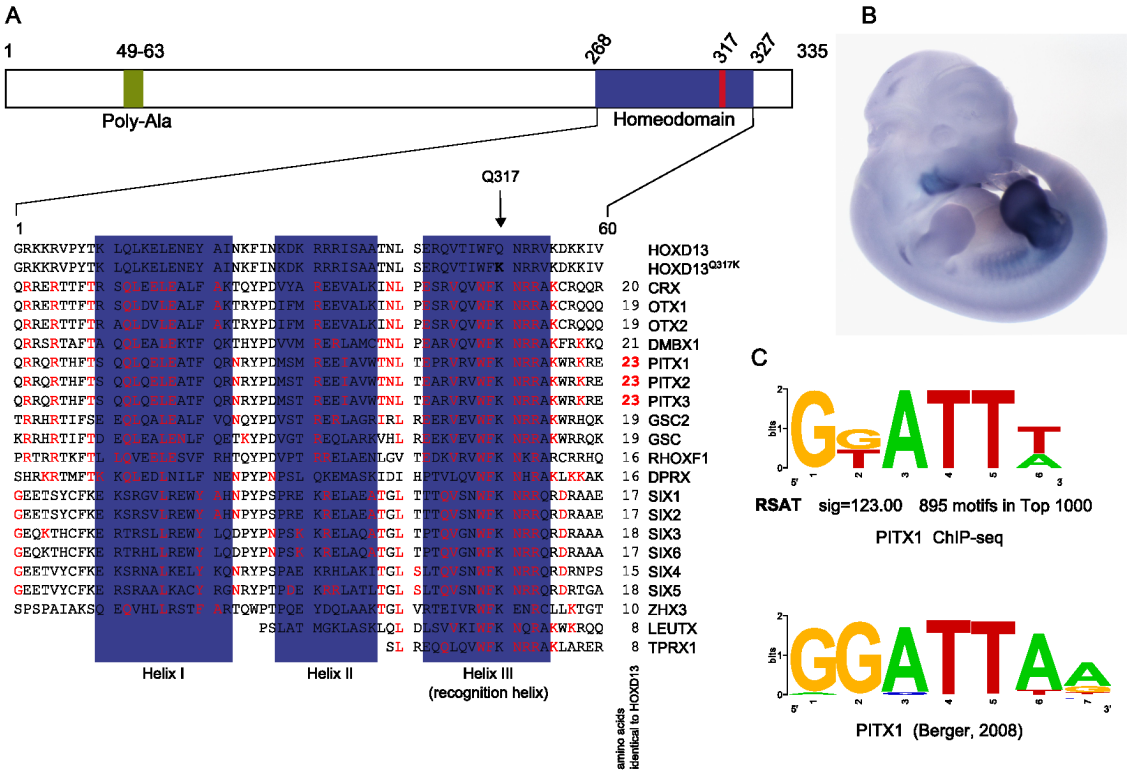


Figure 3.10 The Q317K mutation changes the HOXD13 homeodomain to a bicoid-type homeodomain.

(A) Localization of the Q317K mutation in the HOXD13 protein. Alignment between the human HOXD13 homeodomain sequence and all bicoid-type homeodomains in mouse (Uniprot). All *identical* positions are highlighted in red. (B) Expression of *Pitx1* transcript in E11.5 mouse embryos visualized by *in situ* hybridization (with kind permission from M. Spielmann and D. Lupiañez) (C) *De novo* motif analysis using sequences for the 1000 most significant PITX1 peaks (summit +/- 75bp) as inferred by RSAT.

Taken together, the altered binding specificity of the HOXD13^{Q317} mutants in combination with the importance of PITX1 in limb development raised the possibility, that the pathogenicity of the *HOXD13*^{Q317K} mutation might at least partially be due to aberrant transactivation of genes that are regulated by *PITX1*.

3.3.2 *In vitro* Binding Specificities of HOXD13^{Q317}-mutant Homeodomains

To regulate putative PITX1 target genes, HOXD13^{Q317K} needs to be able to bind to the PITX1 recognition sequence. Since *de novo* motif analysis revealed a very similar, but not identical motif to the known PITX1 recognition sequence, the binding activity of wildtype and Q317 mutant homeodomains in respect to the PITX1 recognition sequence was tested.

For this, EMSAs with oligonucleotides carrying PITX1 and HOXD13 binding sites were conducted. HOXD13^{wt} specifically bound the oligo containing a HOXD13 binding site, but HOXD13^{Q317K} did not show appreciable binding activity (Figure 3.3B and Figure 3.11A). In contrast, the HOXD13^{Q317R} showed considerable residual binding to the wildtype HOXD13 site, reflective of the *de novo* motif analysis, where a HOXD13^{wt}-like motif was detected as secondary motif for HOXD13^{Q317R} but not for HOXD13^{Q317K}.

When testing the PITX1 binding site, wildtype and Q317-mutant homeodomains showed a directly opposite effect. Only the Q317K mutant homeodomain was able to efficiently bind the oligo harboring a PITX1 binding site, whereas the wildtype and Q317R homeodomains were not (Figure 3.11B).

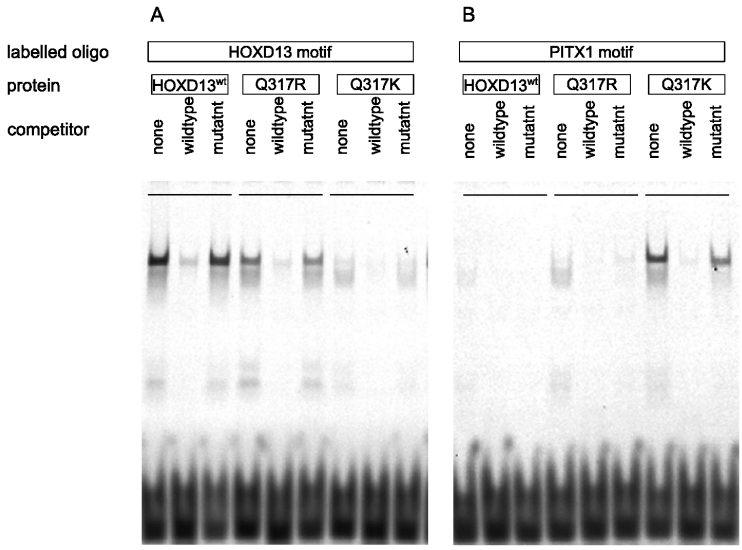


Figure 3.11 HOXD13^{Q317K} , but not HOXD13^{Q317R} binds the PITX1 bindng site *in vitro*

In vitro binding of wildtype and mutant HOXD13 homeodomains to HOXD13 and PITX1 binding sites, respectively. EMSAs were performed as in Figure 3.3 using oligonucleotides containing a HOXD13 (CCAATAAAA) or PITX1 recognition sequences (GGATTA). (A) HOXD13^{wt} and HOXD13^{Q317R} recognize the HOXD13 binding site, whereas HOXD13^{Q317K} does not. (B) Conversely, only the HOXD13^{Q317K} homeodomain binds to the PITX1 binding site, whereas it is not bound by HOXD13^{wt} or HOXD13^{Q317R} homeodomains.

3.3.3 Comparison of *In Vivo* Binding to PITX1

Motif analyses and EMSAs for HOXD13^{Q317K} and HOXD13^{Q317R} raise the question to which extent the genomic binding of the HOXD13 mutants reflected the genomic binding of either HOXD13 or PITX1.

To identify PITX1 binding sites in the chicken genome, ChIP-seq from chMM-cultures infected with *PITX1*-expressing RCASBP-virus was performed and identified 40,881 binding sites for PITX1. *De novo* motif analysis of ChIP-seq peaks readily identified the known PITX1 motif (Figure 3.11B), indicating functionality of the protein. Next, with the genomic data generated for HOXD13^{wt}, HOXD13^{Q317R}, HOXD13^{Q317K}, and PITX1 at hand, the genome-wide binding of the four factors was compared using three separate methods.

I. Principle Component Analysis (PCA)

ChIP-seq experiments produced genome-wide coverage profiles that were characteristic and reproducible for each factor. To compare the genome-wide binding in an as unbiased way as possible, Principle Component Analysis (PCA) was selected to assess the similarity of the coverage profiles. The read count distribution across the genome in 500bp windows for the

eight datasets, representing the replicates for HOXD13^{wt}, HOXD13^{Q317K}, HOXD13^{Q317R} and PITX1, were used as input. The PCA result, when presented in a biplot as shown in Figure 3.12A, clusters data points the closer together the more similar the datasets are. As expected, the biological replicates for each condition clustered closely together. The data points for HOXD13^{Q317K} were located at an intermediate position between those of HOXD13^{wt} and PITX1 (Figure 3.12A), indicating a shift in the genomic binding of HOXD13^{Q317K} towards that of PITX1. In contrast, the data points for HOXD13^{Q317R} remained closer to those of HOXD13^{wt}.

II. Peak Overlap

As mentioned above, HOXD13^{wt} bound 34,267 sites, HOXD13^{Q317K} 25,036, HOXD13^{Q317R} 21,396, and PITX1 40,881. Based on a restrictive definition of shared binding events (the summits of overlapping peaks must be within 100bp of on another), most peaks were found to be TF specific and only 45 sites were bound by all four factors. However, HOXD13^{Q317K} and HOXD13^{Q317R} showed striking differences when comparing the overlap with HOXD13^{wt} or PITX1 peaks. HOXD13^{Q317R} shared 27.5% (5887) of its binding sites with HOXD13^{wt}, whereas HOXD13^{Q317K} shared only 6.0% (1492) (Figure 3.12B). In contrast, HOXD13^{Q317R} shared just 2.9% (633) of its binding sites with PITX1 but HOXD13^{Q317K} 6.6% (1671). The wildtype HOXD13 also shared only 1.1% (383) of its binding sites with PITX1 (Figure 3.12B). Additionally, HOXD13^{Q317K} and HOXD13^{Q317R} shared a large set of target sites (3094) that were bound by neither HOXD13^{wt} nor PITX1.

III. Read Count Distribution

Weak binding events are signified by weak read enrichments and are not always detected as peaks. In turn, this might lead to an underrepresentation of shared binding events based on the sole observation of overlapping peaks. To take this into account, the *mean read density* from HOXD13^{wt}, HOXD13^{Q317K}, HOXD13^{Q317R}, and PITX1 ChIP-seq experiments surrounding the identified binding sites for each factor was determined using seqMINER (Figure 3.12C). As expected, a strong read enrichment for each factor was detected surrounding its peaks. There was no read enrichment in HOXD13^{wt} or HOXD13^{Q317R} ChIP-seq experiments at the position of PITX1 peaks. In contrast to HOXD13^{wt} and HOXD13^{Q317R}, there was a moderate read enrichment at the PITX1 peaks in the HOXD13^{Q317K} ChIP-seq, indicating HOXD13^{Q317K} binding at PITX1 peaks. HOXD13^{Q317K} reads also displayed a moderate enrichment at HOXD13^{wt} peaks and somewhat stronger around HOXD13^{Q317R} peaks. Conversely, read enrichment in the PITX1 ChIP-seq was observed at HOXD13^{Q317K} peaks, but not at HOXD13^{wt} or HOXD13^{Q317R} peaks. Again, the genomic binding of the HOXD13^{Q317K} mutant specifically shifted towards PITX1 sites to an extent that was not observed for HOXD13^{Q317R} or HOXD13^{wt}.

Taken together, all three analyses identified an altered genomic binding of the HOXD13^{Q317K} mutant that shifts to adopt a more PITX1-like binding profile, which is not observed for the HOXD13^{Q317R} mutant.

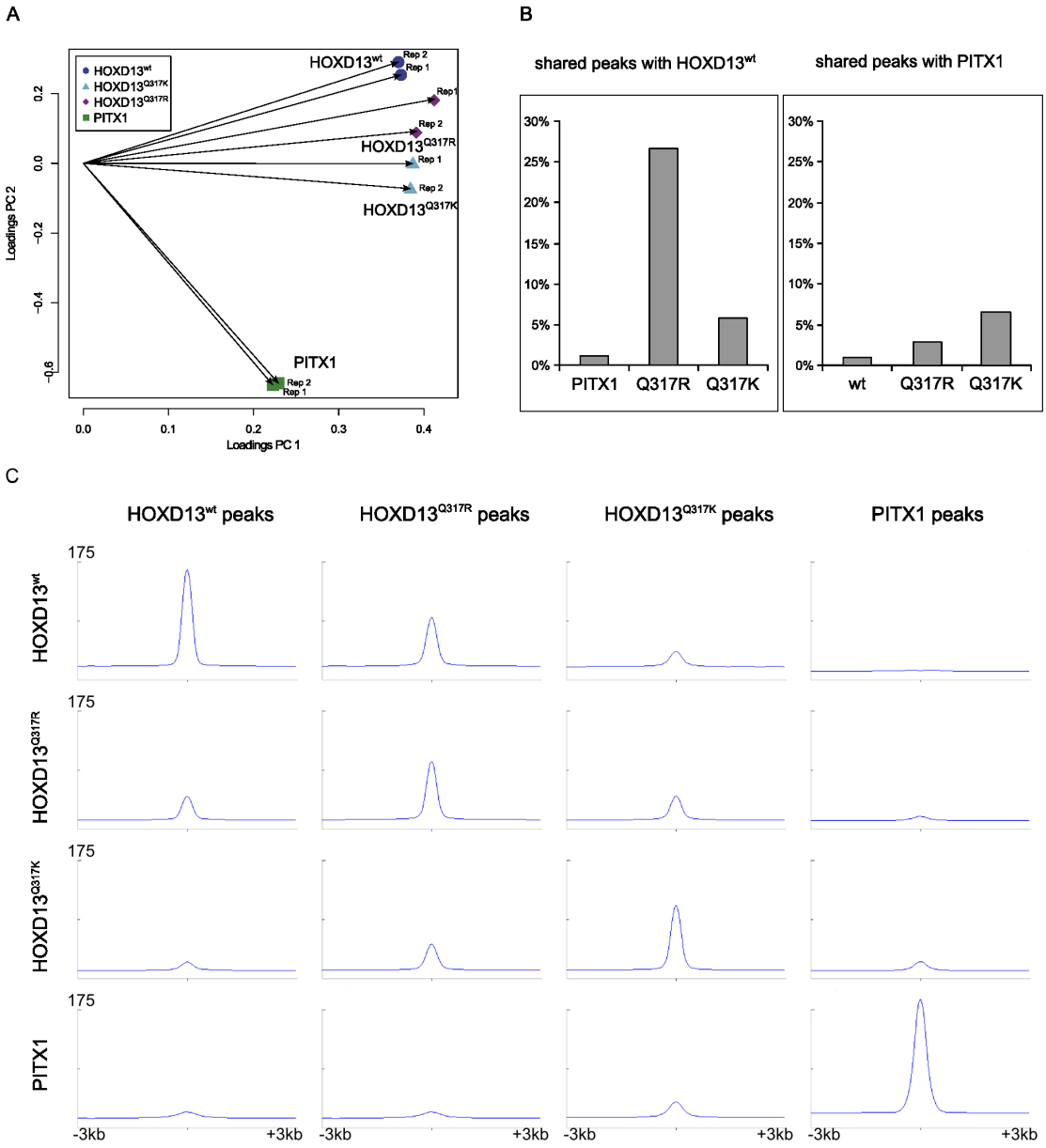


Figure 3.12 The genomic binding of HOXD13^{Q317K} shifts towards PITX1

(A) Principle component analysis for the coverage profiles of the uniquely mapped reads for all eight datasets (two biological replicates each for HOXD13^{wt}, HOXD13^{Q317K}, HOXD13^{Q317R}, and PITX1. (B) Peak overlap with respect to HOXD13^{wt} and PITX1 peaks. (left) PITX1 shares 1.0% (383/40,881) of its peaks with HOXD13^{wt}, HOXD13Q317R 27.5% (5,887/21,396) and HOXD13^{Q317K} 6.0% (1,492/25,036). (right) HOXD13wt shares 1.1% (383/24,267) of its binding sites with PITX1, HOXD13^{Q317R} 3.0% (633/21,296) and HOXD13^{Q317K} 6.7% (1,671/25,036). (C) HOXD13^{Q317K} shares binding properties with both HOXD13^{wt} and PITX1. Mean read densities centred (+/- 3kb) around the peak summits for HOXD13^{wt}, HOXD13^{Q317R}, HOXD13^{Q317K}, and PITX1 (left to right). HOXD13^{wt} (164.1 tags/50bp; 93.5 tags/50bp; 43.2 tags/50bp; 23.6 tags/50bp), HOXD13^{Q317R} (49.7 tags/50bp; 101.2 tags/50bp; 49.8 tags/50bp; 19.5 tags/50bp), HOXD13^{Q317K} (24.9 tags/50bp; 51.9 tags/50bp; 109.1 tags/50bp; 25.6 tags/50bp) and PITX1 (15.1 tags/50bp; 24.2 tags/50bp; 38.9 tags/50bp; 191.2 tags/50bp).

3.3.4 Gene Regulation: HOXD13^{Q317K} and PITX1 Induce Similar Gene Expression

To analyze whether the shift in genomic binding profiles was also reflected in gene regulation, RNA-seq expression analysis for chMM cultures infected with *Hoxd13*^{wt}, *Hoxd13*^{Q317K}, *Hoxd13*^{Q317R}, or *PITX1* was performed. First, differentially expressed genes were identified by filtering expression levels according to a minimal expression of 2 RPKM in at least one experiment and an at least two-fold differential expression in any of the cultures as compared to mock infected cultures. This resulted in 3118 transcripts belonging to 2506 genes that were differentially expressed in at least one of the chMM cultures. Next, to assess general similarities and differences in the induced expression changes, PCA analysis and hierarchical clustering of the expression values for all differentially expressed transcripts were conducted. For the PCA, the resulting biplot demonstrated a similar distribution as seen for ChIP-seq with the HOXD13^{Q317K} data point being close to PITX1 and HOXD13^{Q317R} close to HOXD13^{wt} (Figure 3.13A). Similarly, hierarchical clustering of transcript levels clustered HOXD13^{Q317K} with *PITX1* infected cultures, and HOXD13^{Q317R} with HOXD13^{wt} (Figure 3.13B), indicating that the gene expression changes induced by HOXD13^{Q317K} were more similar to those induced by

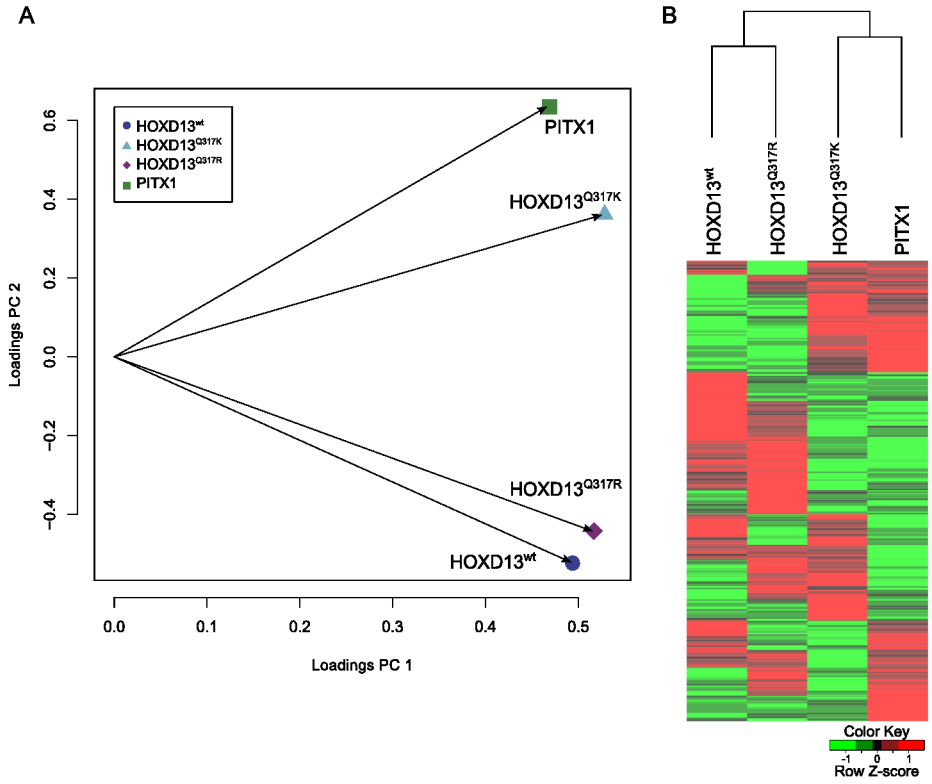


Figure 3.13 Regulation by HOXD13^{Q317K} is more similar to PITX1 than to HOXD13^{wt} or HOXD13^{Q317R}.

Global expression analysis of genes regulated by HOXD13^{wt}, HOXD13^{Q317R}, HOXD13^{Q317K}, or PITX1 in chMM cultures. RNA-seq expression analysis found 3118 transcripts (2506 genes) to be at least two-fold differentially regulated in *Hoxd13*^{wt}, *Hoxd13*^{Q317R}, *Hoxd13*^{Q317K} or *PITX1* chMM compared to mock infected cultures (RPKM cut-off 2). (A) Principle component analysis (PCA) of the expression values (normalized to expression values in mock infected cultures) of these genes shows a similar clustering as seen in the ChIP-seq PCA. (B) Hierarchical Clustering of the expression values (RPKM) clusters *Hoxd13*^{Q317K} with *PITX1* and *Hoxd13*^{Q317R} with *Hoxd13*^{wt}. Red colors indicate up-regulation, green colors down-regulation as compared to mock-infected cultures.

PITX1 than to HOXD13^{wt} or HOXD13^{Q317R}.

For each factor, a comparable number of genes were differentially expressed (HOXD13^{wt} 1006 genes, HOXD13^{Q317R} 845, HOXD13^{Q317K} 1102, PITX1 1014). Interestingly, when comparing the genes regulated by mutant or wildtype HOXD13 with PITX1 regulated genes, 436 genes were co-regulated by HOXD13^{Q317K} and PITX1, whereas only 262 genes were co-regulated by HOXD13^{wt} and PITX1 and 237 genes by HOXD13^{Q317R} and PITX1 (Figure 3.14A).

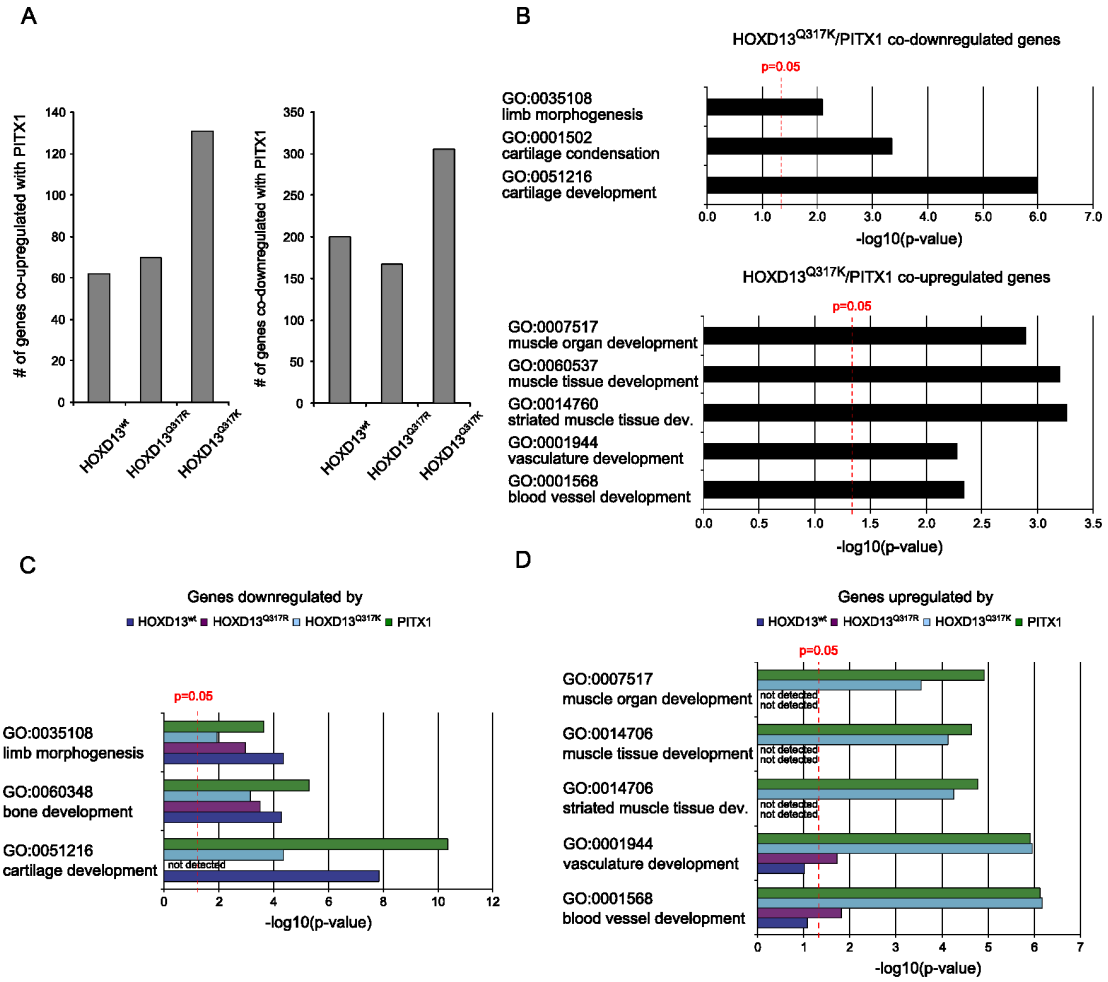


Figure 3.14 HOXD13^{Q317K} and PITX1 activate genes involved in muscle and vasculature development

(A) HOXD13^{Q317K} co-regulates nearly twice as many genes with PITX1 than HOXD13^{wt} or HOXD13^{Q317R}. The number of genes co-up-regulated (left) or co-down-regulated (right) with PITX1 for HOXD13 wildtype and mutants. (B-D) GO analysis of genes differentially regulated by the four factors individually and genes co-regulated by HOXD13^{Q317K} and PITX1. For each factor, a list of differentially regulated genes, compared to mock-infected cultures was generated and used for GO-analysis with DAVID . The bars represent the negative log10 of the p-values of the selected GO terms. (B) Genes associated with limb morphogenesis, cartilage development, and cartilage condensation, were enriched in the set of genes down-regulated by HOXD13^{Q317K} and PITX1 (top). Genes associated with GO terms related to muscle development and vasculature development, were enriched in the set of genes up-regulated by HOXD13^{Q317K} and PITX1 (bottom). (C) Genes associated with GO terms limb morphogenesis, bone development and cartilage development were enriched in the set of down-regulated genes for all four factors (except chondrocyte differentiation for HOXD13^{Q317R}). The colored bars represent the negative log10 of the p-value found for selected GO terms found in each of the sets of down-regulated genes (D) In the set of up-regulated genes, genes associated with GO terms related to muscle differentiation and vasculature development were strongly enriched among the up-regulated genes in HOXD13^{Q317K} and PITX1 chMM cultures, but not for HOXD13^{wt} and HOXD13^{Q317R}. The colored bars represent the negative log10 of the p-value for selected GO terms found in each of the sets of up-regulated genes.

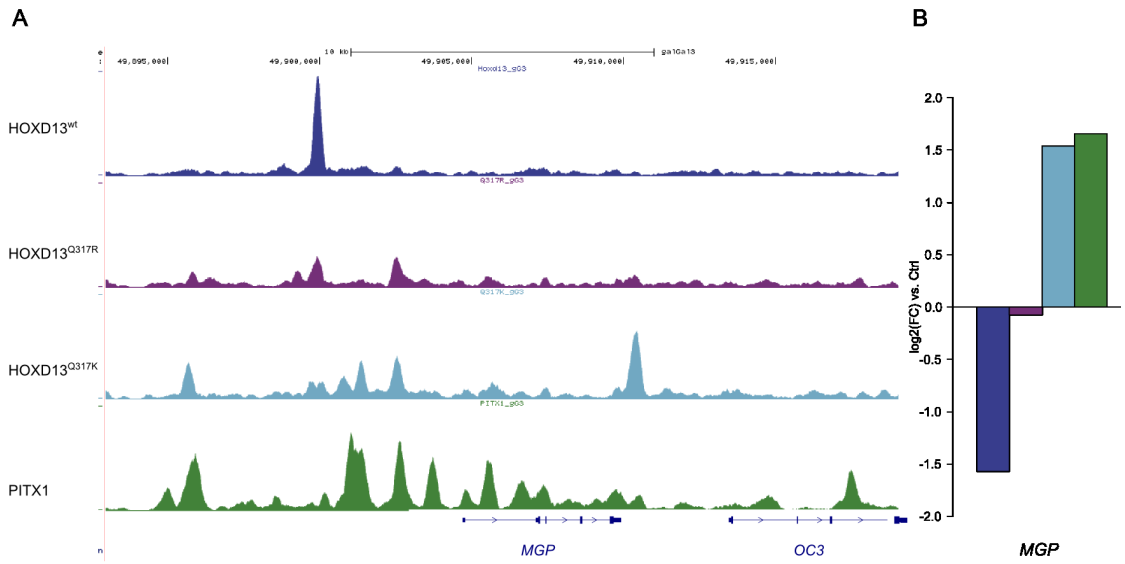


Figure 3.15 Binding profiles at the gene locus of the differentially regulated *MGP* gene

(A) Coverage profiles for HOXD13^{wt}, HOXD13^{Q317R}, HOXD13^{Q317K} and PITX1 in a 20 kb region covering the *MGP* gene show the similarity in binding between HOXD13^{Q317K} and PITX1. (B) *MGP* is 3-fold up-regulated in *PITX1* and *Hoxd13^{Q317K}* infected chMM cultures and 3-f^^old down-regulated by *Hoxd13^{wt}* cultures (Log2 fold changes compared to mock-infected cultures).

Next, to identify those among the 436 HOXD13^{Q317K}/PITX1 co-regulated genes that might contribute to the similarity found in PCA and hierarchical clustering analysis, a Gene Ontology (GO) analysis on the 305 co-down-regulated and 131 co-up-regulated genes was performed. GO terms related to cartilage differentiation and limb development were overrepresented in the set of co-down-regulated genes, which were also found to be overrepresented in the sets of down-regulated genes for each individual factor (Figure 3.14B,C). In contrast, genes annotated to several GO categories related to muscle development as well as blood vessel development were overrepresented among the HOXD13^{Q317K}/PITX1 co-up-regulated genes and, intriguingly, these and related GO terms were also enriched among HOXD13^{Q317K} and PITX1 up-regulated genes and were not similarly enriched in the HOXD13^{wt} or HOXD13^{Q317R} up- or down-regulated genes (Figure 3.14).

3.3.5 Differential Binding in the Vicinity of Co-Regulated Genes

Genome-wide comparison of TF-binding demonstrated that HOXD13^{Q317K} specifically binds to many PITX1 binding sites and that a group of target genes is co-regulated by only HOXD13^{Q317K} and PITX1 as compared to HOXD13^{wt} and HOXD13^{Q317R}. The next analysis attempted to identify whether co-regulation and cobinding are associated.

An intriguing example for this hypothesis is depicted in Figure 3.15A, which shows the binding profiles of the four TFs at the genomic region surrounding the *Matrix-Gla protein* (*MGP*) gene. *MGP* encodes the Matrix-Gla protein, which inhibits calcification in proliferating chondrocytes

and blood vessels (Luo et al. 1997; Yao et al. 2011). HOXD13^{wt}, HOXD13^{Q317K} and PITX1 bind to different sites in the upstream region of *MGP*, whereas HOXD13^{Q317R} shows only weak binding in this region. HOXD13^{wt} binds a unique site, while HOXD13^{Q317K} and PITX1 share three binding sites with additional unique binding sites for HOXD13^{Q317K} and PITX1 each. The *MGP* gene is also among the set of co-regulated genes, as it is 3-fold up-regulated in *Hoxd13^{Q317K}* and *PITX1* infected cultures, but 3-fold down-regulated in *Hoxd13^{wt}* infected cultures and not regulated in *Hoxd13^{Q317R}* cultures (Figure 3.15B).

In order to systematically test whether HOXD13^{Q317K}/PITX1 co-regulated genes are also cobound by the factors, the genes had to be identified as cobound or not cobound. For this, all genes were classified as cobound when a HOXD13^{Q317K}/PITX1 shared peak was found in the gene body or within 25kb up- or downstream of the gene. Using this classification, 13.1% (57) of the 436 co-regulated genes were also cobound by HOXD13^{Q317K}/PITX1, significantly more than expected by chance (Fisher's exact p=1.295x10^{-3}) (Figure 3.16). As a control, an equivalent analysis performed for the HOXD13^{Q317R} mutant and PITX1 and found a non-significant 5.06% (12 out of 237) of co-regulated genes cobound (p=0.1335) (Figure 3.16).

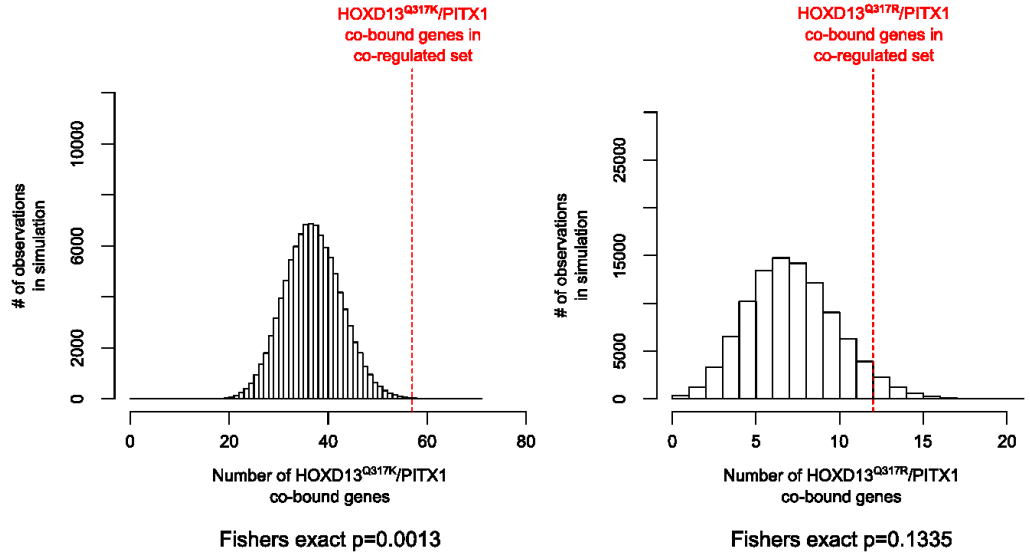


Figure 3.16 HOXD13^{Q317K} and PITX1 share binding sites in the vicinity of co-regulated genes

More HOXD13^{Q317K}/PITX1 co-regulated genes were significantly more co-bound (13.1% (57 of 436); red line) than expected by chance (Fishers exact p=1.295x10^{-3}). The histogram shows the distribution of co-bound genes when randomly selecting 346 genes (100,000 iterations). 5.1% (12 of 237) of HOXD13^{Q317R}/PITX1 co-regulated genes were also co-bound (red line), not significant (Fishers exact p=0.1335). The histogram shows the distribution of co-bound gene numbers when randomly selecting 237 genes (100,000 iterations).

3.3.6 *Hoxd13^{Q317K}* and *PITX1* Induce Similar Phenotypic Effects

Chicken micromass cultures are derived from limb bud mesenchymal cells. Cultured at high density, they spontaneously differentiate into various cell lineages including chondrocytes which form cartilage nodules that undergo differentiation, mineralization and finally bone

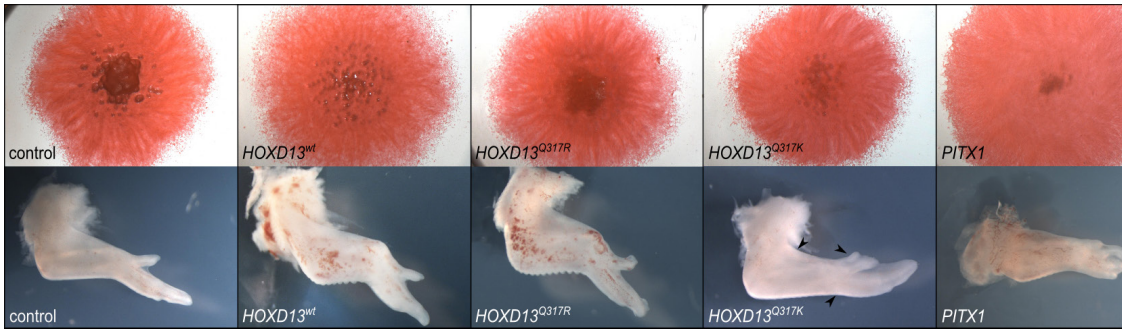


Figure 3.17 Some effects on cell differentiation and development are shared between *Hoxd13*^{Q317K} and *PITX1*.
(Upper row) Eosin-stained chMM cultures for *Hoxd13*^{wt}, *Hoxd13*^{Q317R}, *Hoxd13*^{Q317K}, and *PITX1*. Both *Hoxd13*^{Q317K} and *PITX1* induced differentiation into thick, fibrotic tissue (day 9). (Lower row) Viral overexpression of *Hoxd13*^{wt}, *Hoxd13*^{Q317R}, *Hoxd13*^{Q317K} and *PITX1* in chicken wing buds (HH36). *Hoxd13*^{wt} and *Hoxd13*^{Q317R} overexpression leads to an overall shortening of skeletal elements. *Hoxd13*^{Q317K} infected wing buds lack or have a reduced forelimb-specific patagium, have a straightened wrist and develop a fourth digit (arrowheads). *PITX1* overexpression leads to a partial hind- to forelimb transformation.

formation (Ahrens et al. 1979). Viral expression of wildtype *Hoxd13* strongly inhibits chondrogenic differentiation. Instead, cells in every part of the culture except the very center do not assemble into organized structures (Figure 3.17). In contrast, chMM cultures overexpressing *Hoxd13*^{Q317R}, although a little smaller in size, appear to undergo the same chondrogenic differentiation process as mock infected chMM (Figure 3.17). ChMM infection with *Hoxd13*^{Q317K} resulted in a completely different phenotype. The inhibition of chondrogenic differentiation was even stronger, but in contrast to *Hoxd13*^{wt}, the cells formed a multi-layered fibrotic tissue, radially extending from the center of the culture (Figure 3.17). This very characteristic growth of the culture was also observed in *PITX1*-expressing chMM cultures.

So far, the experiments demonstrated that binding specificity, gene expression and morphology of *Hoxd13*^{Q317K} infected chMM cultures shift from *Hoxd13* to *PITX1* characteristics. To test whether this shift in specificity can also be observed *in vivo*, *Hoxd13*^{wt}, *Hoxd13*^{Q317R}, *Hoxd13*^{Q317K}, and *PITX1* were expressed in chicken limb buds using the same 3xFLAG-tagged constructs as used for ChIP-seq experiments. The effects of *Hoxd13*^{wt} and *PITX1* in this experiment have been described and are clearly distinguishable. On the one hand, expression of *PITX1* in the developing chick wing bud led to a partial fore-to-hindlimb transformation of the infected wing (Logan and Tabin 1999; Szeto et al. 1999). The infected wings showed an absence of the patagium (the skin that will form the surface of the wing), demonstrated straightening of the posterior wrist flexure, and developed an additional digit (Figure 3.17). On the other hand, ectopic expression of *Hoxd13*^{wt} led to an overall reduction in wing size but had no effect on limb patterning (Figure 3.17) (Goff and Tabin 1997).
Like *Hoxd13*^{wt}, viral infection of wing buds with *Hoxd13*^{Q317R} mainly led to an overall reduction of the limb structures (Figure 3.17). In a minority of cases, however, reduced interdigital cell death at the first and between the first and second digit, as well as mispositioning of the wrist could be observed.

In strong contrast, *Hoxd13*^{Q317K} infected wing buds repeatedly lacked or had a reduced patagium, and exhibited a straightening of the posterior flexure of the wrist. Moreover, a fourth digit developed; however, the position of this digit was different from that observed in *PITX1* infected wings. This additional digit extended from the first phalanx of digit 1, whereas with *PITX1* overexpression, the additional digit was located completely apart from digit 1 (Figure 3.17 (arrowheads)).

Collectively, the analyses of *Hoxd13*^{Q317K} expression in chMM cultures as well as the effects induced in limb bud injections clearly demonstrate that HOXD13^{Q317K} causes distinct effects from those induced by HOXD13^{wt} or HOXD13^{Q317R} and displays a partial overlap with the effects of wildtype *PITX1*.

3.4 Functional Characterization of the HOXD13^{R298Q} Mutation

Standard biochemical approaches (see Section 3.2) demonstrated a moderate loss-of-function effect for the HOXD13^{R298Q} mutation. The mutant protein was still able to bind the HOXD13 recognition sequence *in vitro* and had a 20% reduced activation capacity in reporter assay (Figure 3.3). In contrast, the mutation causes a fully penetrant SPD phenotype and a not fully penetrant BDA2 phenotype in patients, whereas *HOXD13* haploinsufficiency only leads to SPD with low penetrance (Goodman et al. 1998; Goodman and Scambler 2001).

Therefore, the chMM-ChIP-seq system was applied to further elucidate the binding behavior of HOXD13^{R298Q}. First, the chondrogenic differentiation of *3xFLAGHoxd13^{R298Q}* and *3xFLAGHoxd13^{R298W}* infected chMM cultures was compared to *3xFLAG-Hoxd13^{wt}* and to mock infected chMM cultures. As expected, *Hoxd13^{wt}* strongly inhibited chondrogenic differentiation. Alcian Blue staining of the *Hoxd13^{R298Q}* infected cultures, however, showed a reduced inhibition of chondrogenic differentiation, whereas the *Hoxd13^{R298Q}* infected cultures showed no effect on differentiation at all (Figure 3.18).

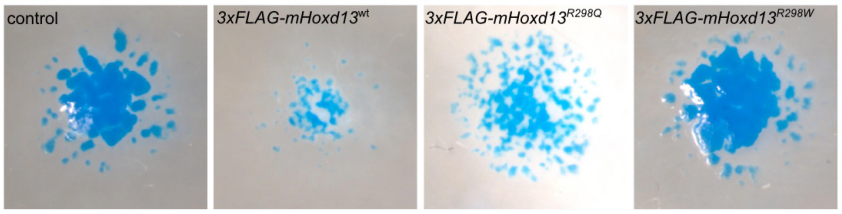


Figure 3.18 HOXD13^{R298Q} inhibits the chondrogenesis less than HOXD13^{wt}

Alcian Blue stained chMM cultures (day 9) infected with the same amount empty, *3xFLAG-Hoxd13^{wt}*, *3xFLAG-Hoxd13^{R298Q}*, and *3xFLAG-Hoxd13^{R298W}* expressing RCASBP(A) virus.

3.4.1 Binding Specificity

Next, chromatin from *Hoxd13^{R298Q}*-infected chMM cultures was used for ChIP-seq. Two biological replicate experiments were subjected to quality control, tested for reproducibility (IDR), and 21,347 reproducible HOXD13^{R298Q} binding sites were identified (for details, see Section 5).

Analogous to the initial analysis of Q317 mutants, the 1000 most significant binding sites were used for *de novo* motif analysis using DREME and RSAT. The primary HOXD13^{R298Q} motif closely resembled the HOXD13^{wt} motif (Figure 3.19), however, a small difference, an A-to-G substitution at position 5 that is not present in the wildtype (asterisk in Figure 3.19), was observed. Notably, both motif-analysis tools identified a composition of secondary motifs that was unlike HOXD13^{wt}, HOXD13^{Q317K}, or HOXD13^{Q317R}, where the majority of secondary motifs resembled the primary motif.

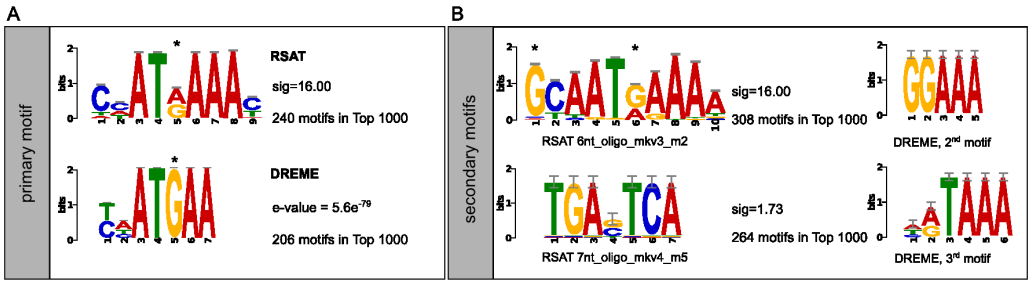


Figure 3.19 HOXD13^{R298Q} binds a site similar to HOXD13^{wt}

De novo motif analysis using sequences for the 1000 most significant HOXD13^{R298Q} peaks (summit +/- 75bp) as inferred by RSAT and DREME. (A) The HOXD13^{R298Q} mutation (top) recognizes motif similar to HOXD13^{wt} (Figure 3.8). The major difference is at position 5, where HOXD13^{R298Q} prefers a G (DREME) or an A or a G (RSAT). Sig and e-values indicate the significance of the motif. The numbers refer to the number of motifs found in the Top1000 peaks (FIMO p <0.0001). (B) Selected secondary motifs from RSAT (left) show (top) a refined version of the primary motif that shows an unusual G at position 1 and (bottom) a HOX-unrelated motif. Secondary motifs found with DREME (right) for HOXD13^{R298Q} show a HOX-unrelated GGAAA and the an [A/G]TAAA motif that resembles the HOX9-13 core motif. Motifs were drawn using the MEME suite (Bailey et al. 2009).

RSAT detected the non-HOX motif (TGA[G/C]TCA) (Figure 3.19B) among the secondary motifs and DREME found the unrelated GGAAA and [A/T/C][A/G]TAAA at 2nd and 3rd position, respectively. However, to a lower extent secondary motifs resembling the primary were also detected for HOXD13^{R298Q}, as can be seen for the refined primary motif at 2nd position (RSAT, Figure 3.19B). Intriguingly, although motif analyses for all factors were performed on the 1000 most significant peaks for each factor, the significance for the primary motif for HOXD13^{R298Q} was the lowest of all HOXD13 variants (compare e-values and sig-values of Figure 3.19 to Figure 3.8 and Figure 3.9). In line with this observation, when using FIMO to identify individual motif occurrences, HOXD13^{R298Q} also was the HOXD13 variant with the least primary motifs in its Top 1000 peaks (DREME: 206/RSAT: 240).

Taken together, initial analysis of the HOXD13^{R298Q} ChIP-seq data demonstrates that the mutant locates to the nucleus and binds DNA-sequences that are highly similar to the wildtype recognition sequences, although they are present in fewer HOXD13^{R298Q} binding sites. Furthermore, the data also indicate that the composition of binding sites seems to be more diverse than for HOXD13^{wt} or the HOXD13^{Q317} mutants

3.4.2 Binding Intensity

The results of the initial motif analysis indicated that the molecular effect of the R298Q substitution might be more subtle than the switch in sequence specificity seen for the Q317 mutations. Moreover, there are several pathogenic mutations in homeodomain proteins that alter the arginine affected by the R298Q mutation (R31, within the homeodomain), most importantly a *HOXD13^{R298W}* mutation associated with a mild form of SPD (Debeer et al. 2002). In that case, the authors proposed loss of DNA-binding as underlying pathomechanism, because a R31H

substitution in MSX2 ($MSX2^{R31H}$) did not change the sequence specificity, but rather reduced the binding affinity of a mutant MSX2 homeodomain (Wilkie et al. 2000).

Therefore, to assess the DNA-binding affinity of wildtype and R298-mutant homeodomains, quantitative EMSAs were performed. Different concentrations (5, 10, 15, and 20ng) of purified $HOXD13^{wt}$, $HOXD13^{R298Q}$, and $HOXD13^{R298W}$ homeodomains were mixed with labeled oligonucleotides harboring the HOXD13 recognition sequence. At comparable protein concentrations, DNA-binding by $HOXD13^{wt}$ homeodomains was strongest, that of $HOXD13^{R298Q}$ homeodomains clearly weaker and that of $HOXD13^{R298W}$ mutant homeodomain was very strongly reduced (Figure 3.20).

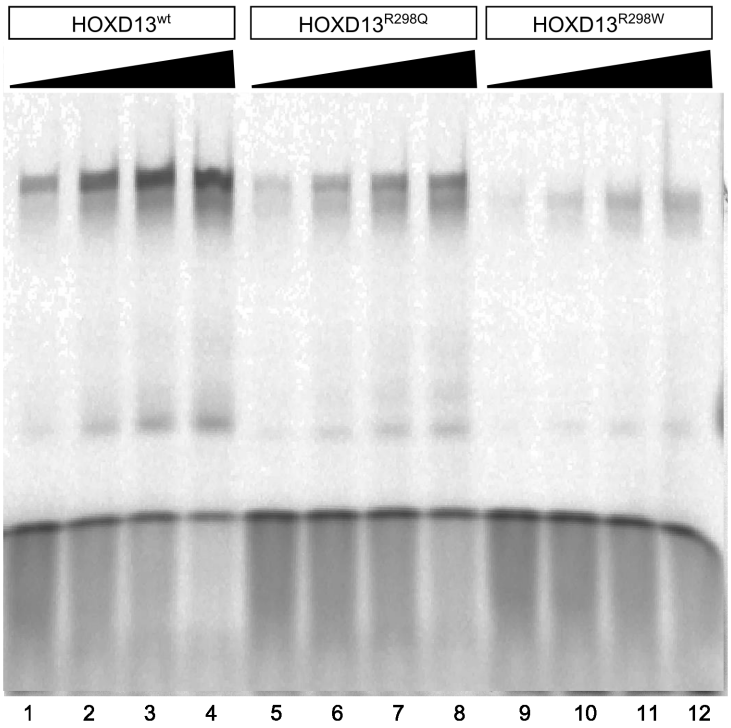


Figure 3.20 DNA-binding affinity in $HOXD13^{R298}$ mutants is reduced

In vitro binding affinity of wildtype and R298-mutant HOXD13 homeodomains were measured in gel-shift assays (EMSA). Fluorescent dye labeled double-stranded oligonucleotides containing a binding site (CCAATAAAA) for HOXD13 were incubated with different concentrations (5, 10, 15, and 20ng) of purified wildtype and mutant homeodomains. The wildtype HOXD13 displayed stronger retention of oligonucleotides at lower concentrations indicating higher binding affinity. $HOXD13^{R298Q}$ bound slightly lower, $HOXD13^{R298W}$ with weak affinity.

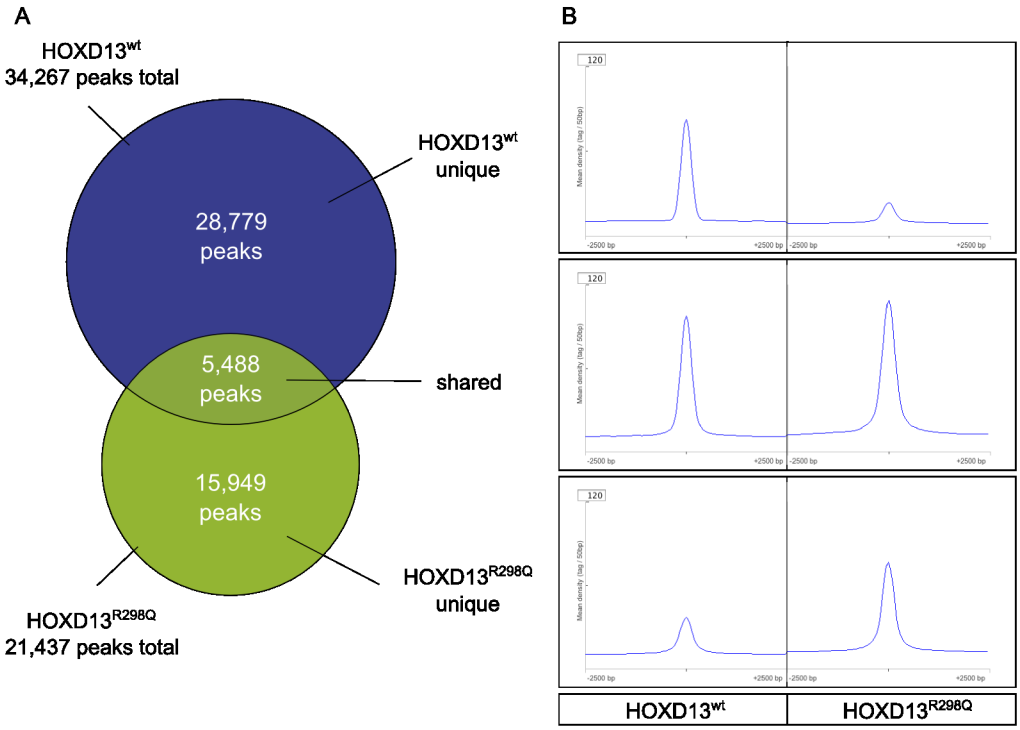


Figure 3.21 HOXD13^{wt} and HOXD13^{R298Q} weakly bind each others unique peaks

(A) Overlap of HOXD13^{wt} and HOXD13^{R298Q} binding sites found in ChIP-seq. Calculation of the overlapping peaks is described in Materials and Methods. (B) Mean read enrichment in HOXD13^{wt} and HOXD13^{R298Q} ChIP-seq surrounding (+/-2500bp) the HOXD13^{wt}-unique, shared and HOXD13^{R298Q}-unique peaks. Residual enrichment of HOXD13^{R298Q} at HOXD13^{wt}-unique peaks is weaker than HOXD13^{wt} enrichment at HOXD13^{R298Q}-unique peaks.

3.4.3 Differential Motif Analysis

Motif analysis and gel shift assays found only minor differences between HOXD13^{wt} and HOXD13^{R298Q}. Therefore, the ChIP-seq data was then analyzed to see whether the genomic binding of HOXD13^{R298Q} is similar or distinct from HOXD13^{w}

First, the overlap between the binding sites was calculated. Strikingly, despite similar sequence specificity of HOXD13^{wt} and HOXD13^{R298Q}, only 5488 peaks (25.7% of all HOXD13^{R298Q} peaks) were bound by both factors. Consequently, this leaves 28,779 HOXD13^{wt}-unique and 15,949 HOXD13^{R298Q}-unique peaks (Figure 3.21).

To further profile the genomic binding surrounding the HOXD13^{wt}-unique, shared, and HOXD13^{R298Q}-unique peaks, seqMINER was used to display the read distribution both ChIP-seq experiments. As expected, the *mean read density* (enrichment) surrounding the HOXD13^{wt}-unique peaks was higher for HOXD13^{wt} than for HOXD13^{R298Q} and the same was true *vice versa*. Also, reads were strongly enriched at the shared peaks in both datasets (Figure 3.21B).

However, there was a notable difference in enrichment at the HOXD13^{wt} and HOXD13^{R298Q}-unique peaks. At the HOXD13^{wt}-unique peaks, the mean enrichment by HOXD13^{wt} was 3.5-

fold stronger than the HOXD13^{R298Q}-enrichment. On the other hand, at the HOXD13^{R298Q}-unique peaks, the mean enrichment of the mutant was only 2-fold stronger than HOXD13^{wt}-enrichment (Figure 3.21).

In addition to displaying the *mean read density* surrounding provided peak lists, the seqMINER program also groups the peaksets into subclusters of peaks that are bound in a similar manner or intensity. The program clustered HOXD13^{wt}-unique, shared, and HOXD13^{R298Q}-unique peaks into ten subclusters with similar binding modes. Then, the ten subclusters were grouped by hand into three clusters with similar binding intensities (Figure 3.22). Strikingly, this clustering revealed that the read enrichment at subclusters is more variable than suggested by the *mean read density* presented in Figure 3.21 (compare to Figure 3.22A).
For example, one subcluster of HOXD13^{wt}-unique peaks (HOXD13^{wt} very strong) was very strongly enriched in the HOXD13^{wt}-ChIP-seq but only very weakly in the HOXD13^{R298Q}-ChIP-seq (4- to 5-fold difference) (Figure 3.22A, top). Another group of HOXD13^{wt} -unique peaks (HOXD13^{wt}-moderate) was moderately enriched by HOXD13^{wt} and weakly by HOXD13^{R298Q} (3.3-fold difference). Conversely, the HOXD13^{R298Q}-unique peaks did also form a number of subclusters. However, these HOXD13^{R298Q}-unique subclusters all showed stronger HOXD13^{R298Q} enrichment by but did not differ as starkly in residual HOXD13^{wt} enrichment (Figure 3.22A, bottom). The ratio between HOXD13^{wt}-enrichment and HOXD13^{R298Q}-enrichment was similar (2.1- to 2.4-fold difference) between all groups, indicating uniformly stronger HOXD13^{R298Q} binding. Again, the residual HOXD13^{wt} enrichment at each group of HOXD13^{R298Q}-unique peaks was higher than it was for HOXD13^{R298Q} at HOXD13^{wt}-unique peaks[7].
The most dramatic differences between subclusters were observed for the HOXD13^{wt}/HOXD13^{R298Q} shared peaks (Figure 3.22A, middle). The shared peaks – independently detected as bound in both ChIP-seq experiments – consist of three different subclusters; HOXD13^{wt}-stronger (505), equally bound (4027) and HOXD13^{R298Q} stronger (956) peaks.

Next, to identify differences between the various subclusters, sequences from each set were subjected to *de novo* motif analysis using RSAT. The primary and two secondary motifs detected for each peakset of Figure 3.22A are depicted in Figure 3.22B. Intriguingly, a systematic difference in sequence composition between predominantly HOXD13^{wt}-bound and predominantly HOXD13^{R298Q}-bound peaks was revealed when investigating the peak subclusters independently (Figure 3.22B).

[7] There was one subcluster of HOXD13^{R298Q}-unique peaks that showed an equally strong enrichment in HOXD13^{wt} and HOXD13^{R298Q} ChIP-seq experiments .This is possibly due to HOXD13^{R298Q} peaks that were not detected by peak calling or were filtered out in the overlap-process.

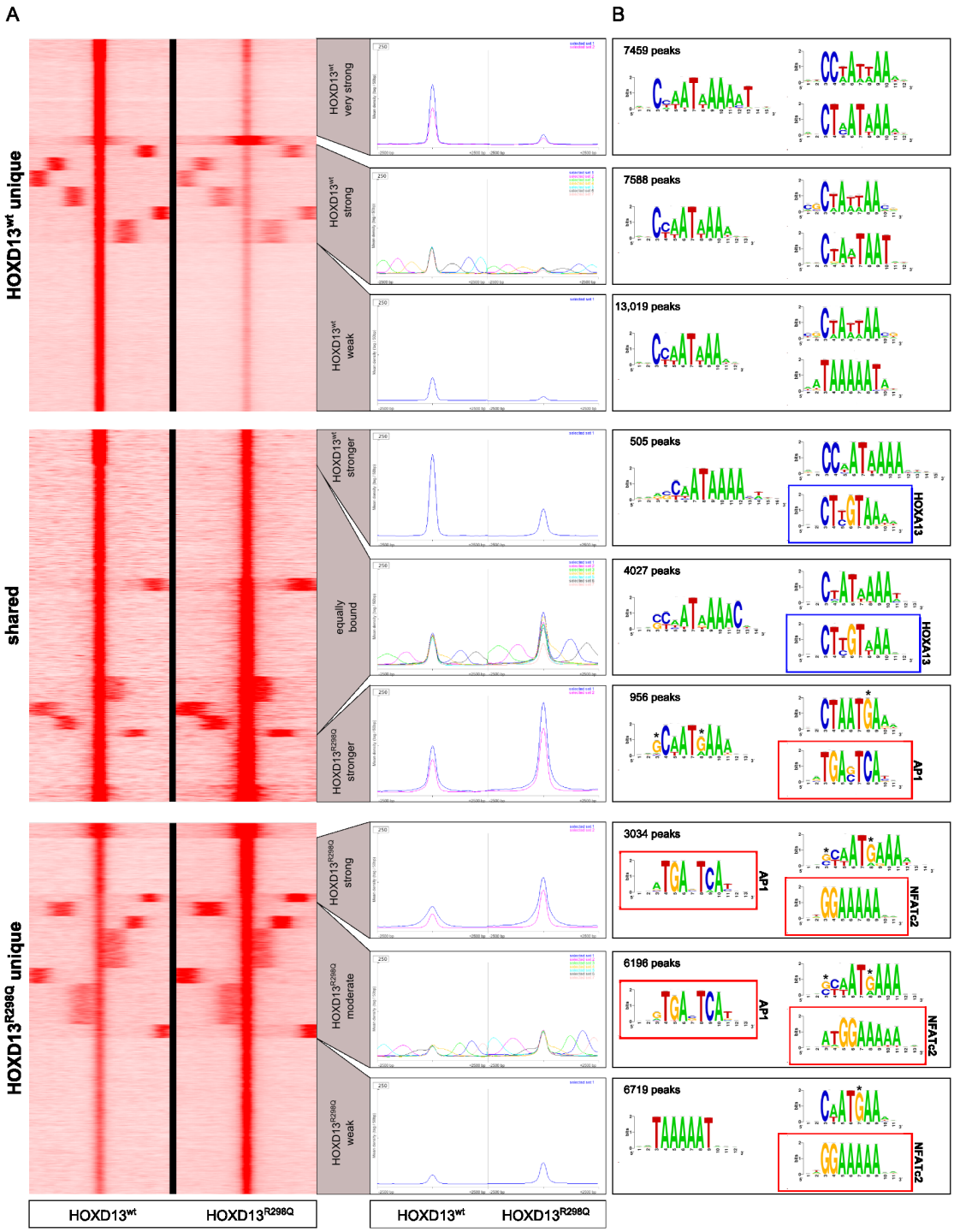


Figure 3.22 Differential motif analysis of HOXD13^{wt}-unique, HOXD13^{R298Q}-unique and shared peaks

(A) Read count distribution (+/- 2500bp) surrounding the peak sets. (left) heatmap demonstrating the different clustered as detected by seqMINER. (right) mean read distribution of the indicated subclusters of peaks. (B) *De novo* motif analysis (RSAT) using the sequences of indicated subclusters as input. Asterisks indicate changes to the HOXD13^{wt} motif, red boxes motifs unique to predominantly HOXD13^{R298Q} bound sites, blue boxes indicate *de novo* detection of HOXA13-like motifs.

On the one hand, a *bona fide* HOXD13 motif was identified in all subclusters that showed stronger read enrichment by HOXD13^{wt}. On the other hand, the motifs identified for predominantly HOXD13^{R298Q}-bound peaks revealed two striking changes. First, similar to the HOXD13^{R298Q} primary motif, the HOXD13-like motif was changed at the 1st and 6th position (Figure 3.22B, asterisks). Second, two HOX-unrelated motifs were identified as overrepresented

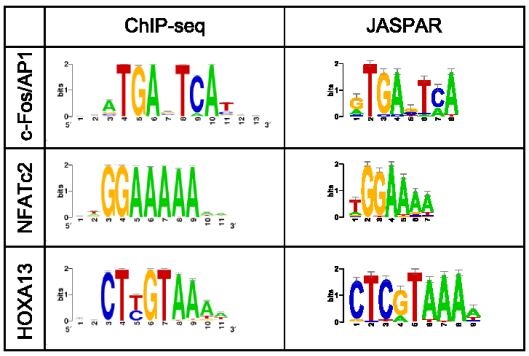


Figure 3.23 Motifs discoverd in differential motif analysis resemble motifs described for other TFs

Motifs found in ChIP-seq (left) compared to highly similar published (JASPAR) motifs (right). Motifs were drawn using the MEME suite (Bailey et al. 2009).

in several subclusters (Figure 3.22B, red boxes). Further, among the motifs identified for the two subclusters of the shared peaks was another HOX-like motif (Figure 3.22, blue boxes). Comparison of these three motifs to the JASPAR database showed their similarity to known motifs. The sTGAsTCAs detected in predominantly HOXD13^{R298Q}-bound peaks is a *bona fide* motif of AP1-transcription factors (Figure 3.23). The other motif detected as secondary motif only in HOXD13^{R298Q}-unique peaks and not in HOXD13^{wt}-unique or shared peaks was GGAAAA that closely resembles the motif of the NFATc2 TF (Figure 3.23). Finally, the HOX-like motif overrepresented in the shared peaks is highly similar to the known HOXA13-motif (Figure 3.23).

The differential motif analysis of HOXD13^{wt} and HOXD13^{R298Q} peaks subclusters highlighted three differences between wildtype and mutant. First, predominantly HOXD13^{R298Q}-bound motifs differ slightly but reproducibly from the HOXD13^{wt}-motif. Second, two HOX-unrelated motifs of potential co-factors are enriched in the predominantly HOXD13^{R298Q}-bound peaks. Finally, the HOXA13 motif appears among the shared peaks that are predominantly bound by HOXD13^{wt}

3.4.4 The Distribution of AP1-Motifs in HOXD13^{wt} and HOXD13^{R298Q} Peaks

The *de novo* motif analyses for the HOXD13^{R298Q} peaks found an enrichment of AP1 recognition sites, which suggested that HOXD13^{R298Q} might directly or indirectly interact with AP1 TFs.

Therefore, AP1 binding site containing peaks were identified using FIMO; 5% (1500) and 13.7% (3111) of HOXD13^{wt} and HOXD13^{R298Q} peaks, respectively, harbored an AP1 binding site (Figure 3.24A). Next, Centrimo analysis was used to profile the distribution of AP1 sites in AP1-containing peaks (Figure 3.24B). There was no central enrichment of AP1-sites in HOXD13^{wt} peaks, but a slight central enrichment of AP1-sites in HOXD13^{R298Q} peaks (Figure 3.24B).

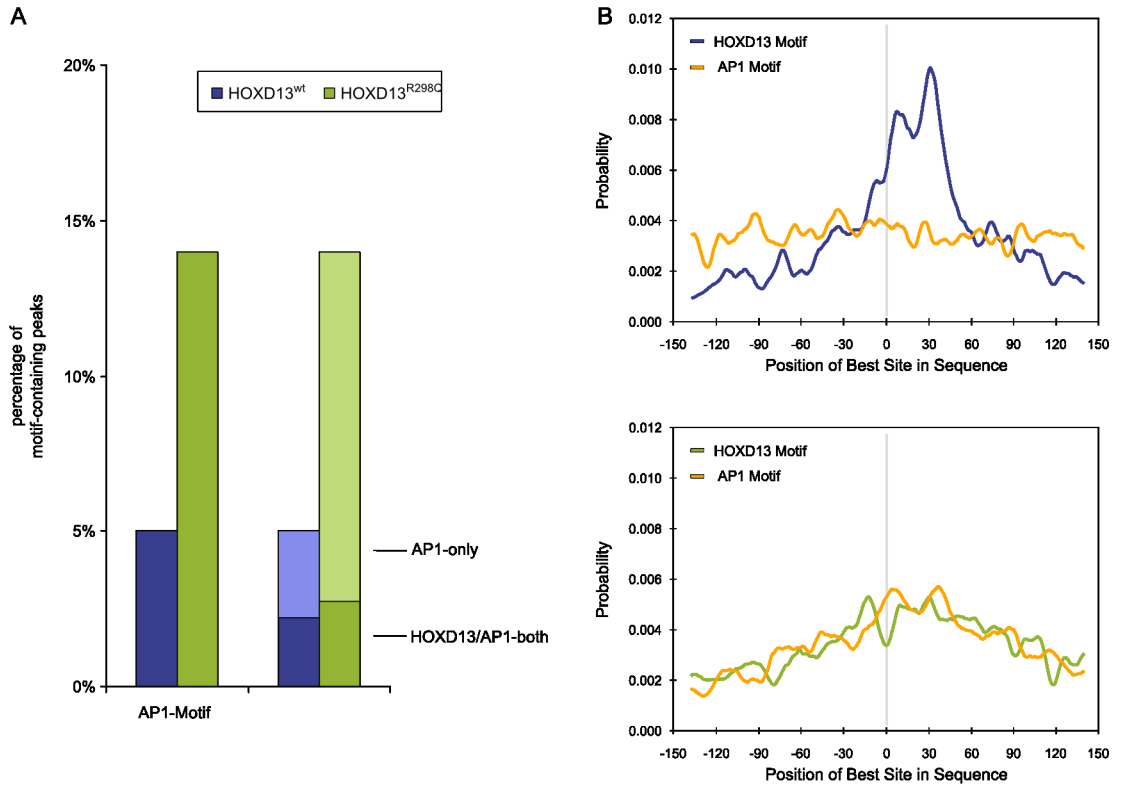
To further profile the distribution of AP1-sites, the peaks were divided into peaks where FIMO identified both HOXD13 and AP1 binding sites (HOXD13/AP1-both) and into peaks that only had an AP1 binding site (AP1-only). For the 1500 AP1-containing HOXD13^{wt} peaks, half were HOXD13/AP1-both and half were AP1-only peaks. Of the 3111 HOXD13^{R298Q} peaks, only 20% were HOXD13/AP1-both, whereas 80% were AP1-only peaks (Figure 3.24A).

The 80% HOXD13/AP1-both peaks for HOXD13^{R298Q} indicated that the mutant might recognize these peaks via tethered binding to an AP1 TF. If this were the case, the AP1 binding sites in those peaks should be enriched centrally, surrounding the summit of the peaks. Therefore, Centrimo was used to determine the distribution of AP1 motifs in the HOXD13/AP1-both and in the AP1-only peaks (Figure 3.24C). For HOXD13^{wt}, neither the HOXD13/AP1-both nor the AP1-only peaks showed central enrichment of the AP1 motif. In contrast, the HOXD13^{wt} motif was centrally enriched in both peaksets (Figure 3.24C, left). For HOXD13^{R298Q}, both motifs, AP1 and HOXD13, were not centrally enriched in HOXD13/AP1-both peaks (Figure 3.24C, top right). In the AP1-only peaks, the graph describes a curve that might be interpreted as weak enrichment of AP1-like sites (Figure 3.24C, bottom right).

Taken together, FIMO and Centrimo analysis did not reveal discernible central AP1 binding site enrichment in AP1-containing HOXD13^{R298Q} peaks.

Figure 3.24 (next page) Analysis of AP1 binding sites in HOXD13^{wt} and HOXD13^{R298Q} peaks.

(A) Percentage of peaks containing an AP1 binding site (lighter shade are AP1-only, darker shaed HOXD13/AP1-both peaks) as determined by FIMO ($p<0.0001$). (B, C) Centrimo analysis of the best sequence match to the AP1 or HOXD13 motif of all HOXD13^{wt} or HOXD13^{R298Q} peaks (B), or of HOXD13/AP1-both or AP1-only peaksets (C). HOXD13^{wt} peaksets show the best sequence match to a HOXD13 motif centrally, near the peak summit, whereas no positional enrichment of AP1 binding sites is detected. In HOXD13^{R298Q} peaks sets, a minor central HOXD13 site enrichment can be seen in the AP1-only peaks. However, there is just a minor central enrichment for AP1 sites in AP1-only HOXD13^{R298Q} peaks.


A
HOXD13wt
HOXD13R298Q
20%
15%
10%
5%
0%
percentage of motif-containing peaks
AP1-only
HOXD13/AP1-both
AP1-Motif
B
HOXD13 Motif
AP1 Motif
Probability
Position of Best Site in Sequence
C
HOXD13wt
HOXD13R298Q
HOXD13 and AP1 Motif in peak
AP1 Motif only in peak

3.4.5 The Distribution of NFATc2-Motifs in HOXD13^{wt} and HOXD13^{R298Q} Peaks

The second motif identified in predominantly HOXD13^{R298Q} bound peaks was a GGAAAA that has been described as recognition sequence for the Nuclear Factor of Activated T-Cells 2 (NFATc2) transcription factor, which is important in T-Cell development, but also expressed in the cartilage of murine E14.5 embryos (Ranger et al. 2000).

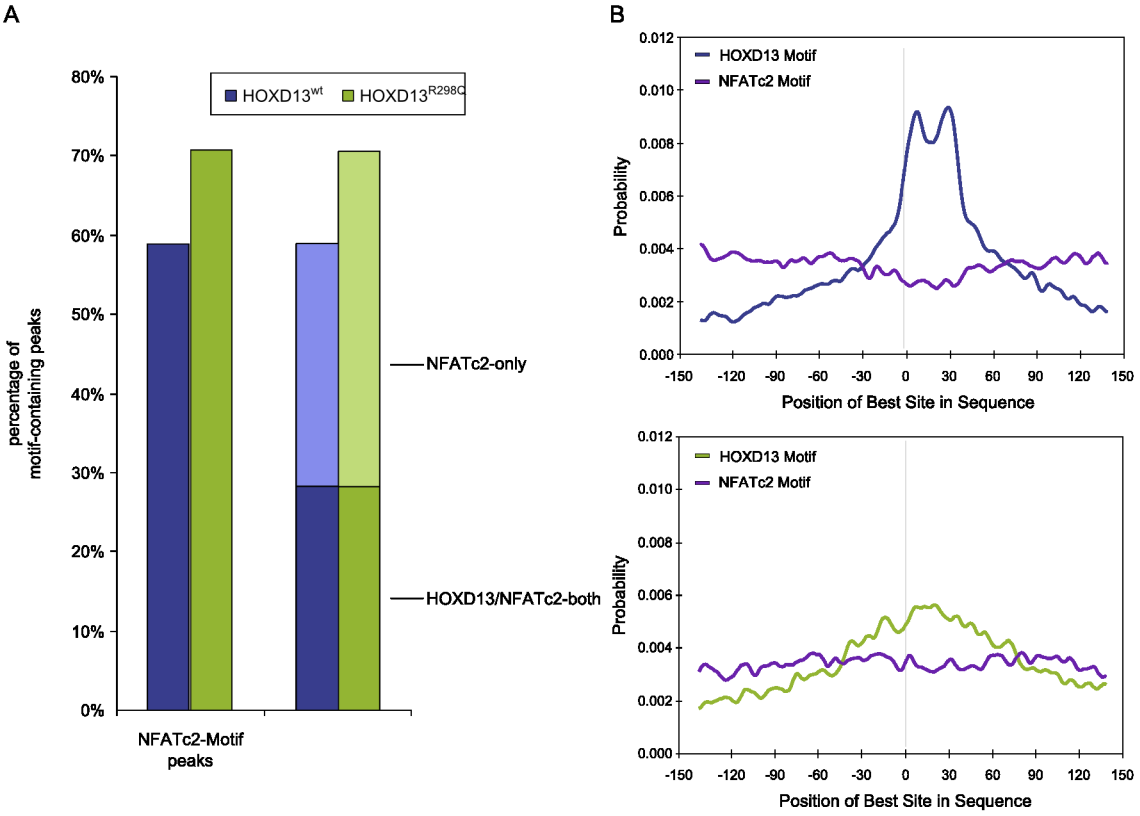
As for AP1, FIMO was used to identify wildtype and mutant peaks containing NFATc2 binding sites. The stringency of the GGAAA motif (large letters mean little variability) prevented identification of putative NFATc2 binding sites at a p-value<0.0001. However, when the cutoff threshold was set to p<0.001 potential binding sites were found in HOXD13^{wt} and HOXD13^{R298Q} peaks (the worst sequence match detected as NFATc2 binding site with p<0.001 was GGAAT). With this lowered cutoff, most HOXD13^{wt} (57%) and even more HOXD13^{R298Q} peaks (71%) were identified as NFATc2-site containing (Figure 3.25A). Then, Centrimo was used to assess the distribution of NFATc2 in HOXD13^{wt} or HOXD13^{R298Q} peaks, but neither HOXD13^{wt} nor HOXD13^{R298Q} peaks showed central enrichment (Figure 3.25B).

Subsequently, peaks were separated into HOXD13/NFATc2-both and NFATc2-only peaks. For HOXD13^{wt}, approximately equal numbers were NFATc2/HOXD13-both and NFATc2-only, whereas for HOXD13^{R298Q}, 40% of peaks were HOXD13/NFATc2-both and 60% were NFATc2-only peaks (Figure 3.25A). Finally, Centrimo was used to test for the distribution of HOXD13 and NFATc2 binding sites in the HOXD13/NFATc2-both and NFATc2-only peaks (Figure 3.25C). As for the combined peakset, HOXD13^{wt} peaks showed only central enrichment of HOXD13 binding sites and no central enrichment for NFATc2 binding sites. For HOXD13^{R298Q} peaks a weak central enrichment of HOXD13 binding sites was detected (as seen for HOXD13^{R298Q} peaks in general), but the NFATc2 binding sites did not show a central enrichment.

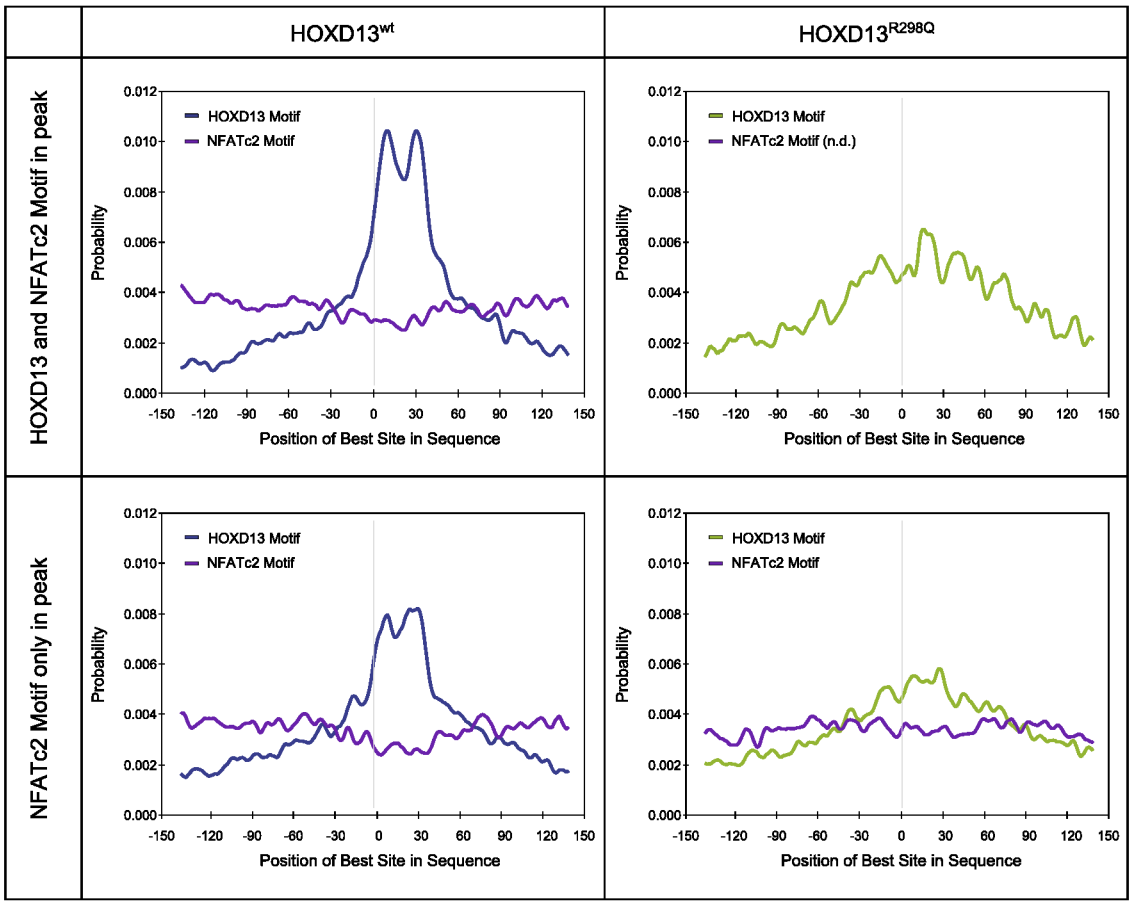
Taken together, the results indicate that the identification of GGAAA sequences in the HOXD13^{R298Q} peaksets might be an artifact and is not related to a gained or lost cofactor interaction of HOXD13^{R298Q}.

Figure 3.25 (next page) Analysis of NFATc2 binding sites in HOXD13^{wt} and HOXD13^{R298Q} peaks.

(A) Percentage of peaks containing an NFATc2 binding site (lighter shade are NFATc2-only, darker shade HOXD13/NFATc2-both) as determined by FIMO (p < 0.001). (B, C) Centrimo analysis of the best sequence match to the NFATc2 or HOXD13 motif in all HOXD13^{wt} or HOXD13^{R298Q} peaks (B), or in HOXD13/NFATc2-both or NFATc2-only peaks (C). All HOXD13^{wt} peaksets show the best sequence match to a HOXD13 motif centrally, near the peak summit while there is no positional enrichment for an NFATc2 motif. In HOXD13^{R298Q} peaksets, the HOXD13 motif shows a slight central enrichment but there is no positional enrichment for potential NFATc2 binding sites.


A
HOXD13wt
HOXD13R298Q
80%
70%
60%
50%
40%
30%
20%
10%
0%
percentage of motif-containing peaks
NFATc2-only
HOXD13/NFATc2-both
NFATc2-Motif peaks
B
HOXD13 Motif
NFATc2 Motif
Probability
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-150 -120 -90 -60 -30 0 30 60 90 120 150
Position of Best Site in Sequence
HOXD13 Motif
NFATc2 Motif
Probability
Position of Best Site in Sequence

C


HOXD13wt
HOXD13R298Q
HOXD13 and NFATc2 Motif in peak
NFATc2 Motif only in peak
HOXD13 Motif
NFATc2 Motif
NFATc2 Motif (n.d.)
Probability
Position of Best Site in Sequence

3.4.6 Distribution of HOXA13 Binding Sites in HOXD13^{wt} and HOXD13^{R298Q} Peaks

General distribution

The differential motif analysis revealed an overrepresentation HOXA13 binding sites among predominantly HOXD13^{wt}-bound peaks, which indicated that HOXD13 might share a significant subset of its binding sites with HOXA13.

To further investigate the relationship between HOXD13 and HOXA13 binding sites, FIMO was used to identify all HOXD13 and HOXA13 binding sites in the HOXD13^{wt} and HOXD13^{R298Q} peaks. Of all HOXD13-peaks, 47.5% (16,279) contained a HOXD13 binding site and 22.5% (7709) contained a HOXA13 binding site (Figure 3.26A). In strong contrast, only 26% of all HOXD13^{R298Q} peaks contained a HOXD13 binding site (FIMO $p<0.0001$, Figure 3.26A) and 11.2% a HOXA13 binding site. To ensure that the lower number of HOXD13 binding sites in HOXD13^{R298Q} peaks was not due to the slightly changed HOXD13^{R298Q} motif, the number of typical HOXD13^{R298Q} binding sites in HOXD13^{wt} and HOXD13^{R29Q} peaks was identified (the motif found in Figure 3.22B, HOXD13^{R298Q}-strong peaks was used for this analysis). In HOXD13^{wt} and HOXD13^{R298Q} peaks, the HOXD13^{R298Q} binding sites were present in only 19% and 22% of peaks, respectively. Collectively, the analysis shows that a large fraction of HOXD13^{wt} peaks contain HOXA13 binding sites, and that HOXD13^{R298Q} peaks generally contain fewer HOX-like binding site compared to HOXD13^{wt} peaks.

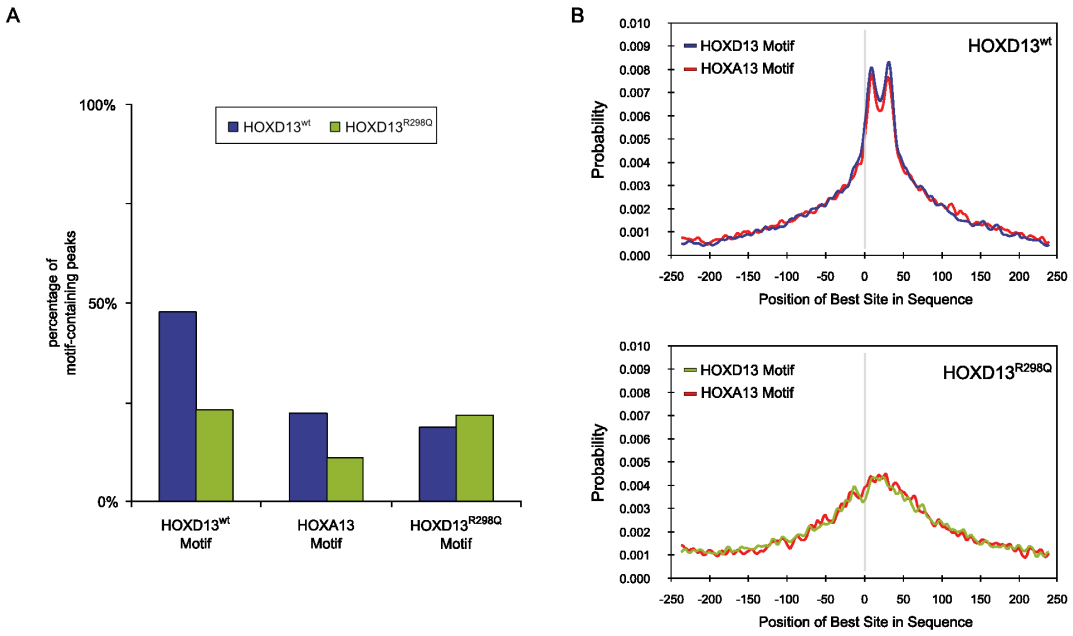
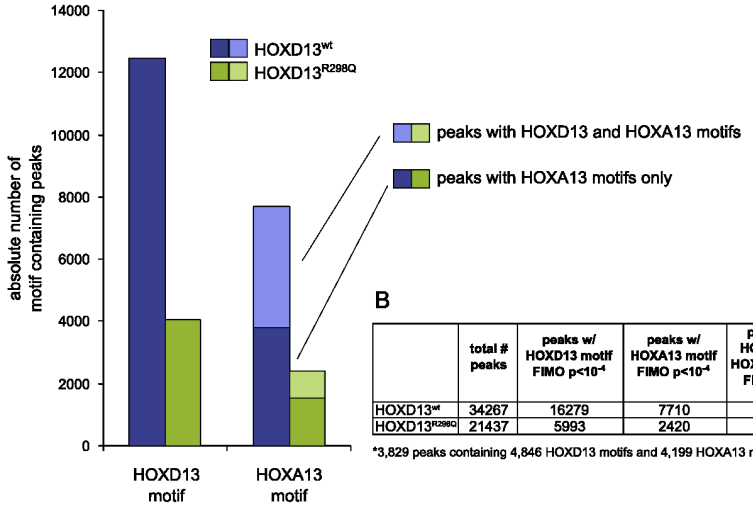


Figure 3.26 Identification of HOXA13 binding sites in HOXD13^{wt} and HOXD13R^{298Q} peaks

(A) Identification of HOXD13^{wt} and HOXD13^{R298Q} peaks containing a HOXD13, HOXA13, and HOXD13^{R298Q} binding sites as identified by FIMO ($p<0.0001$). (B) Distribution of the best sequence match to HOXD13 and HOXA13 motifs in all HOXD13^{wt} (top) or all HOXD13^{R298Q} (bottom) peaks as determined by Centrimo. The HOXD13 and HOXA13 motifs used for this and subsequent analyses are the primary motifs identified by Berger et al. (2008) which were shortened to the central 9bp.

A



B

	total # peaks	peaks w/ HOXD13 motif FIMO $p<10^{-4}$	peaks w/ HOXA13 motif FIMO $p<10^{-4}$	peaks w/ HOXD13 or HOXA13 motif FIMO$<10^{-4}$	peaks w/ HOXD13 motif only	peaks w/ HOXD13 motif only	peaks w/ both motifs
HOXD13^{wt}	34267	16279	7710	20160	12450	3881	3829*
HOXD13^{R298Q}	21437	5993	2420	7774	5353	1780	640

*3,829 peaks containing 4,846 HOXD13 motifs and 4,199 HOXA13 motifs (FIMO $p<10^{-4}$)

Figure 3.27 HOXD13^{R298Q} binds less frequently in sites that contain HOXD13 and HOXA13 binding sites

(A) The absolute number of HOXD13^{wt} and HOXD13^{R298Q} peaks that carry a HOXD13 binding site only and those that carry in HOXA13 binding site. Lighter shaded colors indicate the peaks that contain HOXD13 and HOXA13 binding sites. (B) Tabular description of the data. All binding sites identified by FIMO ($p < 0.0001$).

Following this, the position of the best HOXD13 or HOXA13 recognition sequence within peaks was identified using Centrimo. Usually, for ChIP-seq data, an enrichment of the best sequence match to the motif is centered with a single maximum near the peak summit. Here, a striking distribution of HOXD13 and HOXA13 binding sites within the HOXD13^{wt} peaks was observed. The best sequence match to the HOXD13 (blue) or HOXA13 (red) motif was located at the peak summit or 40 to 50bp downstream of the peak summit, resulting in a graph with two maxima (Figure 3.26B). When performing the corresponding analysis on the HOXD13^{R298Q} peaks, only a weak enrichment of the best sequence matches surrounding the summit was found, and the characteristic profile with two maxima was not seen.

In order to distinguish whether the appearance of HOXA13 binding sites in HOXD13 peaks originates from HOXD13 and HOXA13 binding to separate sequences in the same peak (co-binding) or whether they originate from peaks that only carry a HOXA13 site, the HOXD13 peaks containing a HOXA13 binding site were separated into HOXA13-only or HOXD13/HOXA13-both peaks. The 7709 HOXD13^{wt} peaks containing a HOXA13 binding site split evenly into 3881 HOXA13-only peaks (50.4%) and 3829 HOXD13/HOXA13-both peaks (49.6%) (Figure 3.27). A different distribution was observed for the HOXD13^{R298Q} peaks. Of the 2420 HOXD13^{R298Q} peaks with a HOXA13 binding site, only 545 (22.5%) were HOXD13/HOXA13-both peaks, whereas the large majority (1875/ 77.5%) were HOXA13-only; in other words, the detected HOXA13 binding site was the only HOX-like sequence in the peak. Notably, the majority (356/545) of HOXD13^{R298Q}-peaks containing a HOXD13 and a HOXA13 binding site were shared with HOXD13^{wt}-peaks.

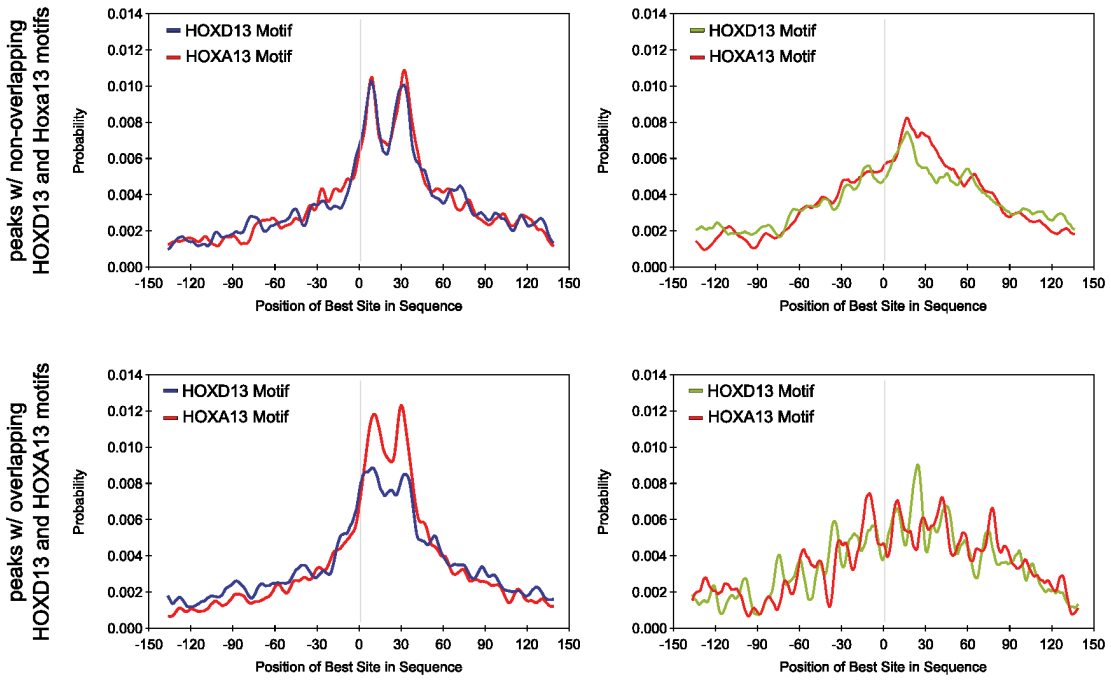


Figure 3.28 Distribution of HOXD13 and HOXA13 motifs at HOXA13 containing peak sets

Distribution of the best sequence match to HOXD13 and HOXA13 motifs in HOXD13^{wt} (left) or HOXD13^{R298Q} (right) peak sets that contain a HOXA13 binding site. Above are the peaks that contain non-overlapping HOXD13 and HOXA13 sites, below are the peaks that only contain a HOXA13 site. Graphs generated using Centrimo.

Centrimo analysis for HOXA13-only and HOXD13/HOXA13-both peaks produced graphs in line with those in Figure 3.26. The characteristic "double-peak" for HOXD13^{wt} was not observed for HOXD13^{R298Q} (Figure 3.28)

Collectively, comparison of HOXA13 binding sites in HOXD13^{wt} and HOXD13^{R298Q} peaks revealed three key observations. First, HOXD13^{R298Q} peaks generally have fewer HOX-like motifs. Second, a surprisingly large number of HOXD13^{wt} peaks contain HOXD13 and HOXA13 binding sites and third, the characteristic co-occurrence of HOXD13 and HOXA13 binding sites is strongly reduced in HOXD13^{R298Q} peaks.

HOXD13/HOXA13 Motif Co-occurrences

The large fraction of HOXD13/HOXA13-both peaks in addition with the intriguing distribution of HOXD13 and HOXA13 binding sites indicated that these peaks might be bound in parallel by HOXD13 and HOXA13 proteins. To see whether the co-occurrence of HOXD13 and HOXA13 binding sites followed any rules that indicate stereotypic binding of HOXD13 and HOXA13, the HOXD13/HOXA13-both peaks were further investigated.

First, the positions of HOXD13 (4846) and HOXA13 (4198) binding sites (FIMO $p<0.0001$) in the 3829 HOXD13/HOXA13-both peaks were identified. Then, for each HOXD13 site, the location of the nearest non-overlapping HOXA13 binding site was determined. This automatically excluded all CTCATAAAA sequences, which are detected by both given HOXD13 and HOXA13 motifs. In total, 3150 HOXD13 binding sites – 15.5% of *all* HOXD13

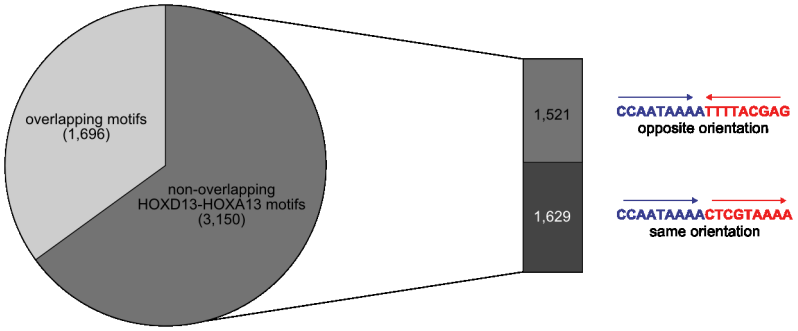
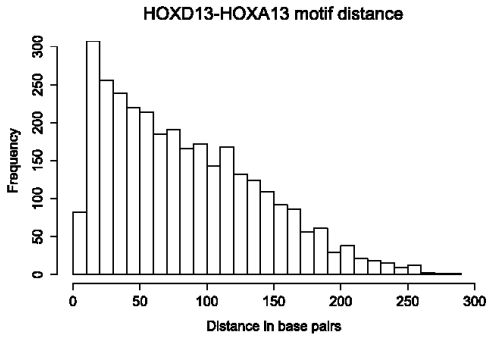


Figure 3.29 Frequent co-occurrence of HOXD13 and HOXA13 binding sites

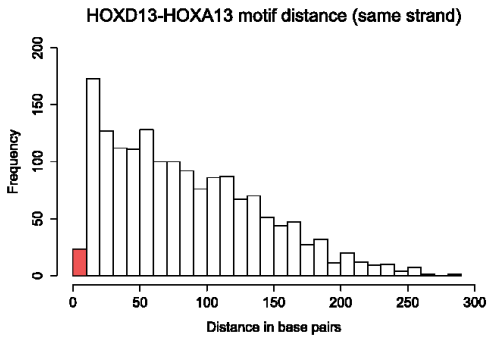
The great majority of HOXD13 and HOXA13 binding sites in HOXD13^{wt} peaks that contain a HOXA13 binding site are non-overlapping. 3,150 HOXD13 binding sites have a second HOXA13 binding site within 300bp. This amounts to 15.5% of all HOXD13 binding sites found in peaks. There is no preferential orientation of the two binding sites towards one another.

binding sites found in peaks – have a second HOXA13 binding site within 300bp radius (Figure 3.29).

Next, to identify preferred distances between the two binding sites, the distance between each binding site pair was calculated. Most HOXA13 binding sites occurred within 40bp of the HOXD13 site (median = 40bp), gradually declining with distance (Figure 3.30). Further, a similar distribution in site distances was observed when considering binding site pairs on the



orientation	total # of peaks	mean	25% quartile	median	75% quartile
any	3150	83.88	36	74	122
same strand	1629	84.27	37	74	122
opposite strand	1521	83.46	35	74	123



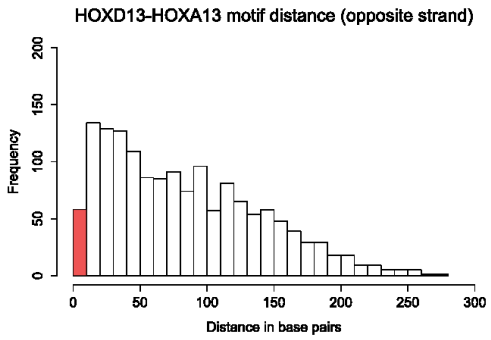


Figure 3.30 Distribution of HOXD13-HOXA13 binding site distances in HOXD13^{wt} peaks

Histogram of distances between HOXD13 and HOXA13 binding sites in 10 bp steps. Most binding sites are within 40bp of one another. Binding sites separated by one or two basepairs are more frequently located on opposite strands than on the same strand (red bars).

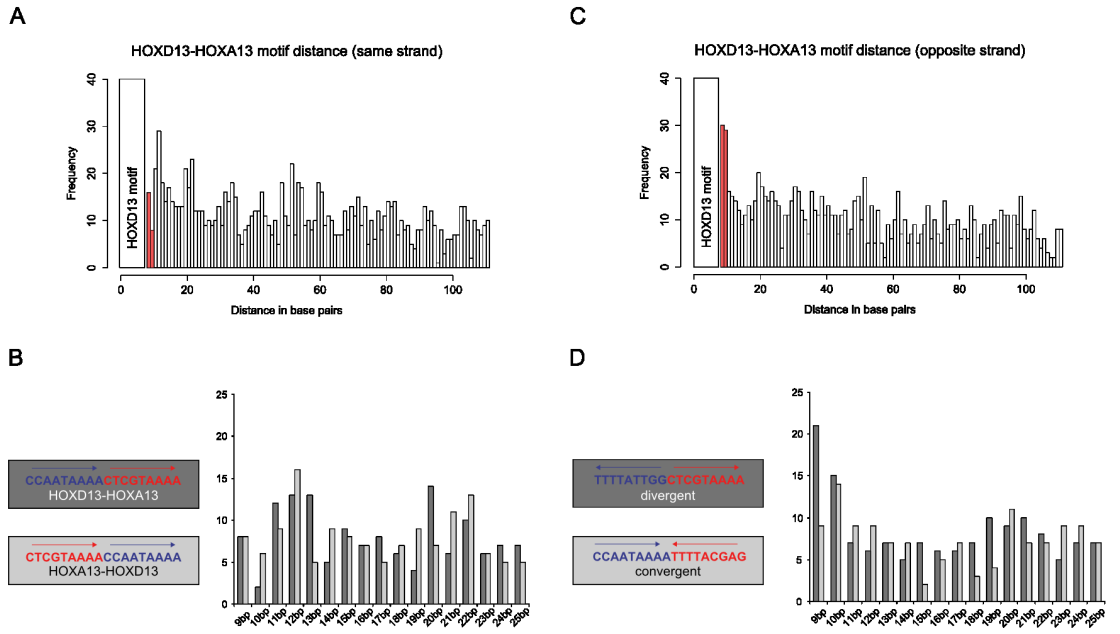


Figure 3.31 Distribution of HOXD13 HOXA13 binding site distances in HOXD13^{wt} peaks

(A, C) Histogram of distances between HOXD13 and HOXA13 binding sites at basepair resolution. Binding sites separated by one or two basepairs are twice as frequently located on opposite strands compared to those located on the same strand (red bars). (B) Histogram of binding site distances (basepair resolution) located on the same strand split into HOXD13-HOXA13 or HOXA13-HOXD13 order. Both orders are equally common. (D) Histogram of binding site distances (basepair resolution) split into divergent and convergent orientation toward one another. When directly adjacent and at 19bp distance the majority of binding sites are oriented divergent to one another.

same and on opposite strands separately, and there was also no preferred orientation of binding sites to one another (1629 same strand, 1521 opposite strand) (Figure 3.29 and Figure 3.30).

However, a striking difference between both orientations could be observed when comparing co-occurring motifs in the same orientation with those of opposite orientation. If the two binding sites are directly adjacent they are counted as separated by 9bp and, if there is a one-basepair gap, by 10bp. Only few HOXD13 sites were adjacent to a HOXA13 site when they were in the same orientation (30), whereas when located on opposite strands, there were approximately twice as many (60) (red bars in Figure 3.30 and Figure 3.31). This difference was even more obvious when looking at the distribution of HOXD13-HOXA13 motif pairs in basepair resolution (Figure 3.31A,C).

Even more specific, if two non-palindromic binding sites are located on opposite strands they can be directed in two different orientations towards one another, convergent or divergent (Figure 3.31B,D). To test whether there are preferences for one orientation over another, the HOXA13-HOXD13 pairs were separated into convergent and divergent pairs. Of the 30 directly adjacent HOXA13-HOXD13 motif pairs in HOXD13^{wt} peaks, 21 were divergently oriented and 9 were convergently oriented, whereas the 30 pairs separated by one basepair were distributed evenly between divergent/convergent (15/14). Interestingly, at 19bp difference – one full

rotation of the DNA double helix – the distribution was again shifted (10 divergent, 4 convergent). In comparison, the orientations for most other distances were distributed evenly, as were the orientations for motif pairs on the same strand (HOXD13-HOXA13 or HOXA13-HOXD13; Figure 3.31). However, due to the low number of individual binding site pairs, these observations were not statistically significant.

Taken together, closer inspection of HOXD13-HOXA13 binding site pairs revealed no preferred distance or orientation for the two binding sites; however, intriguing preferences for binding site orientation could be observed in those pairs, where the HOXA13 binding site lies adjacent to the HOXD13 site.

4 Discussion

4.1 Poor Genotype-Phenotype Correlation of HOXD13 Mutations is Due to Mutation Specific Effects

Transcription factors are crucial for the exact control of gene expression in the developing embryo. Most TF proteins specifically bind to DNA sequences via a DNA-binding domain, while different protein domains interact with other TFs or the transcriptional machinery to activate or inhibit the expression of target genes. TFs are frequently involved in the pathogenesis of disease, in particular congenital malformation syndromes (Boyadjiev and Jabs 2000; Vaquerizas et al. 2009). In many instances this is due to haploinsufficiency, i.e. the functional loss of one allele caused by deletions, nonsense mutations, frameshifts or other inactivating variants. In contrast, missense mutations will not lead to haploinsufficiency, but rather alter TF function by affecting protein-protein or protein-DNA interactions and thereby challenge the interpretation of these mutations.

More than 20 distinct mutations in *HOXD13* have been described (reviewed in Brison et al. 2013). The phenotypic spectrum of *HOXD13* mutations, however, is variable and there is no good genotype-phenotype correlation for HOXD13 mutations, leading to the observation that, depending on the exact type of mutation, the hands and feet can be affected in very different ways. In general, *HOXD13* mutations are associated with SPD. However, *HOXD13* mutations have also been shown to cause brachydactylies (type A2, B, and E), syndactyly type V, clinodactyly of the fifth finger and to change the bone formation process (Debeer et al. 2002; Goodman 2002; Caronia et al. 2003; Zhao et al. 2007; Jamsheer et al. 2012; Wang et al. 2012a). More specifically, haploinsufficiencies and polyalanine expansions in *HOXD13* are associated with synpolydactyly (Muragaki et al. 1996; Goodman et al. 1998), whereas amino acid substitutions in the homeodomain have mostly been described in cases with SPD-unrelated phenotypes affecting the distal thumb and the metacarpals (Caronia et al. 2003; Johnson et al. 2003; Zhao et al. 2007; Wang et al. 2012a).

The presence of several distinctive phenotypes suggests mutation specific disease mechanisms for each type of mutation. In the case of HOXD13 polyalanine expansions, the pathomechanism appears to be a combination of loss of function and gain of function caused by protein aggregation and the interaction with other alanine repeat containing TFs (Villavicencio-Lorini et al. 2010). Considering the drastic differences in phenotypic expression for missense mutations, a common pathomechanism for all mutations is a rather unlikely scenario and specific effects, such as alterations in binding capacity or cofactor interactions are more likely to explain this phenomenon.

In this study, ChIP-seq was applied to characterize two missense mutations in *HOXD13* that are associated with distinct phenotypes. The results of this approach provide direct evidence for the

mutation specific effect of missense mutations for HOXD13 protein function and elucidate the molecular pathomechanims underlying two distinct diseases caused by mutations in the same gene.

4.2 The chMM-ChIP-seq System

4.2.1 Investigation of Mutant TFs

To interpret the effect of a mutation, the full spectrum of TF function needs to be taken into account. TFs are proteins of modular nature and mutations located in the DNA-binding domain will likely lead to a different protein dysfunction than a mutation affecting a protein-interaction domain or a nuclear localization sequence. Unlike enzymes, which stereotypically catalyze specific biochemical reactions, the nature of TFs is to positively or negatively influence the transcription of certain genes inside the nucleus. Current techniques to investigate TF mutations mostly test specific aspects of TF function that often provoke yes-or-no answers. For these approaches prior knowledge of individual TF function and a working hypothesis for the effect of the mutation are required.

ChIP-seq has been used to characterize the genomic binding profiles of numerous TFs in many different experimental settings. However, it has not yet been systematically used to investigate the effects of mutations in TFs associated with hereditary diseases. The main roadblocks consisted in the difficulty in distinguishing between wildtype and mutant TFs with a primary antibody and in obtaining sufficient amounts of tissue of a relevant anatomical site and developmental stage for the analysis.

The chMM-ChIP-seq method developed and applied in this thesis is designed to overcome these hurdles and characterize TF mutations as unbiased as possible and in parallel allowing a flexible setup. One advantage of the system is that it profiles wildtype and mutant TF binding in its environment, the nucleus. In addition, application of a primary cell culture system (chMM) generates comparable conditions for wildtype and mutant proteins on the one hand and avoids the artificial situation of immortalized cell culture lines on the other hand. Furthermore, the chMM system is well suited for investigation of *HOXD13* mutations, as the cells originate from the developing limb bud, where the gene is physiologically expressed. Finally, chMM cultures, in contrast to other cell culture systems, undergo a differentiation process, and thus are well suited to investigate the factors that influence this process.

To date, standard approaches used to elucidate the molecular underpinnings of TF mutations not only have more artificial nature than chMM-ChIP-seq, but also are also limited in the interpretation of the results. Common practices include transient expression of wildtype and mutant proteins followed by immunocytochemistry to test whether folding or nuclear localization of the protein is affected. More specifically, the “activating” ability of a mutant is

assessed in luciferase reporter assays using known target promoters, ideally, but not always, in multiple cell lines. Another common method, if the recognition sequence is known, are EMSAs, which determine the binding affinity of the purified DNA-binding domains *in vitro*. More elaborate methods, such as yeast-2-hybrid or pull-down experiments, can be used to analyze normal or altered protein-protein interactions and are sometimes adapted to screen for potential binding partners (see De Folter and Immink 2011 for an overview). Finally, sophisticated *in vitro* techniques – SELEX, Protein Binding Microarrays (PBM) and Surface plasmon resonance (SPR/Biacore) – could be used to determine differences in DNA-binding specificity and affinity of wildtype and mutant TFs (e.g. Berger et al. 2008; Siggers et al. 2011; Jolma et al. 2013).

To varying degrees, all of these approaches allow only limited insights and/or are artificial in nature. As could be seen for both HOXD13 mutations, correct nuclear localization does not answer whether the protein is functioning correctly inside the nucleus. Luciferase reporter assays, despite having the advantage of comparing sequence requirements for gene activation, are dependent on transient overexpression of TF proteins. The transient transfection leads to TF protein levels that can be much higher than physiological levels and test the activity on target sequences located on plasmid vectors detached from their nuclear environment. Moreover, these reporter assays often show high variability between cell lines, indicating a dependency on cofactors that are not present in all cell types (Whitfield et al. 2012). Regarding protein-protein interactions, the Yeast-2-hybrid and pulldown assays are artificial, as they test for interaction between isolated protein domains in yeast and are prone to identify false-positive interactions (Deane et al. 2002; Chen et al. 2010). PBM and SELEX-seq are less biased in their outcome and could be easily adapted to compare wildtype and mutant TFs. However, in contrast to ChIP-seq assays, these require purified proteins and test their *in vitro* binding. Commonly, only the DNA-binding domain of the TF is used in these assays, since full-length TFs are frequently difficult to purify and despite precise characterization of DNA-binding preferences, *in vitro* assays lack important information of TF binding in the genomic context.

The chMM-ChIP-seq method allows comparative analysis of wildtype and mutant TF binding throughout the genome. *De novo* motif analysis can be used to detect changed binding specificities, whereas the genomic location of binding sites can indicate putative regulatory effects. Additional bioinformatics analyses enable further investigations regarding altered cofactor binding sites or changed distribution of binding sites in relation to genomic landmarks. Moreover, the chMM-system allows analysis of the regulatory effects between wildtype and mutant TFs either by comparing the effect on chondrogenic differentiation in general or by measuring expression levels of individual target genes (see Section 3.2). Thus – performing a single experiment – chMM-ChIP-seq enables to address multiple aspects of TF function that would require separate *in vitro* experiments using common approaches.

However, the chMM-ChIP-seq method is not without limitations. The RCASBP system leads to overexpression of the transduced TF, which in turn might lead to false positive binding sites, although for the comparison of mutant to wildtype TF this is not likely to be relevant. Moreover, absolute quantification of transcript levels could show that the overexpression levels for *Hoxd13* remain in the range of endogenous expression levels in the developing limb bud. In addition, although the chMM system is well suited for studying TFs that influence chondrogenesis, TFs not involved in this process should be tested adequate avian cell culture systems, such as models for muscle, neuronal, and neural crest cell differentiation (McCobb et al. 1990; Dorman and Johnson 1999; Etchevers 2011). Finally, the RCASBP-virus is limited in insert size, so that cDNA of TFs larger than 2kb will likely not be amenable to expression through RCASBP viruses.

Taken together, the chMM-ChIP-seq system enables a relatively quick functional characterization by comparing wildtype and mutant proteins in the genomic context and reveals not only where the mutant TF fails to bind, but also if the protein changes its binding specificity. The ever-growing repertoire of bioinformatic analysis tools allows addressing specific questions regarding binding affinity and specificity, potential cofactor binding sites, or the location of binding sites near regulated genes. Thus, this initial screening of the mutant TF leads to the development of a mutation-specific hypothesis regarding the underlying pathomechanism, which can then be followed by testing the hypothesis in appropriate and custom-designed functional assays.

4.2.2 Possible Applications for the chMM-ChIP-seq Approach

The versatile experimental setup and the chondrogenic differentiation of the chMM cultures enable the adaption of the chMM-ChIP-seq method to address a range of other biological questions.

Since chMM cultures mimic physiological chondrogenic differentiation, the system is well suited to investigate TFs that influence this process. The generation of ChIP-seq profiles for several key TFs in chondrogenic development such as SOX9, RUNX2, SOX5/6 etc. could be used to map common chondrogenic TF binding sites and characterize important *cis*-regulatory elements, as well as elucidating the regulatory relationship of the TFs among each other. Along these lines, coinfection of chMM cultures with adequate viruses the system would provide a good platform to experimentally address regulatory and protein-protein relationships between (chondrogenic) TFs. Finally, generating or integrating chromatin modifications over the developmental time course of chondrogenic differentiation would allow characterization of the changing genomic landscape from undifferentiated mesenchymal progenitor cells to fully differentiated chondrocytes and give insights how TFs influence this process.

4.3 Analysis of two HOXD13^{Q317} Missense Mutations

4.3.1 HOXD13^{Q317} Mutations Induce Distinct Genomic Binding Shifts

The HOXD13^{Q317K} and HOXD13^{Q317R} mutations lead to distinct and novel patient phenotypes, a complex combination of brachy- and oligodactyly, and syndactyly type V (Zhao et al. 2007), respectively, and thus provide a good opportunity to apply the chMM-ChIP-seq technique. Initial motif analysis demonstrated that the sequence-specificity, but not the general DNA-binding ability was altered and that both mutations specifically changed the 5' end of the recognition sequence albeit in slightly different ways.

The result confirmed previous findings of homeodomain-DNA recognition, which showed that the affected glutamine residue is responsible for sequence specificity (Fraenkel et al. 1998). An identical Q to K substitution was introduced in the *Drosophila* fushi-tarazu (ftz) homeodomain (Percival-Smith et al. 1990a; Schier and Gehring 1992), changing the recognition motif *in vitro* from CCATTA to GGATTA, the motif of the homeodomain-protein bicoid (bcd) that endogenously carries a lysine at this position. The altered regulatory capacity ftz^{Q50K} substitution was investigated in *Drosophila* using reporter constructs (Schier and Gehring 1992; Zhao et al. 2000). Flies transgenic for a *ftz^{Q50K}* allele develop normally and do not show any malformations, although reporter constructs respond to the altered binding of ftz^{Q50K} (Schier and Gehring 1992). Experiments using a *Drosophila* cell culture system were able to show the importance of other sequences inside the homeodomains and sequences outside the homeodomain, which are needed to confer the full regulatory capacity (Zhao et al. 2000). In these experiments, the ftz^{Q50K} protein bound with high affinity to and activated constructs carrying consensus bcd binding sites while failing to activate natural enhancers harboring non-consensus binding sites (Zhao et al. 2000).

Motif analysis and EMSAs demonstrated that a similar switch in sequence specificity occurs in the HOXD13^{Q317K} mutant, which recognized the PITX1 binding site *in vitro*. Of the 25 bicoid-type (K50) homeodomain TFs in the mouse genome (20 in humans), only PITX1 is known to play a major role in limb development. *Pitx1* is required for the formation of hindlimb-specific structures and produces hindlimb-like morphologies when misexpressed in forelimbs (Logan and Tabin 1999; Szeto et al. 1999).

The change in sequence specificity introduced by the HOXD13^{Q317R} mutation is similar, but distinct from HOXD13^{Q317K}, as can be seen in the *de novo* motif analysis and EMSAs. Interestingly, although the preferred motif begins with GT, a HOXD13-like motif appears among the secondary motifs in motif analysis. This residual wildtype specificity is also reflected in EMSAs, where HOXD13^{Q317R} still shows considerable binding to the HOXD13^{wt} sequence, whereas this is completely abolished in HOXD13^{Q317K}.

An advantage of this study over previous reports studying mutant homeodomains (Zhao et al. 2000; Caronia et al. 2003; Zhao et al. 2007) is that ChIP-seq allowed to profile the genome-wide binding of mutant HOXD13 proteins and could provide first direct evidence for a genome-wide shift of the mutant towards a more PITX1-like binding pattern. In contrast to the observations for ftz^{Q50K}, the shift in binding is accompanied by a regulatory switch. RNA-seq expression analysis demonstrated that in comparison with $HOXD13^{wt}$ and $HOXD13^{Q317R}$, $HOXD13^{Q317K}$ leads to a clearly different regulation of target genes, many of which are also recognized and regulated by PITX1 (Figure 3.16). This is likely due to the fact that $HOXD13^{Q317K}$ is capable of binding to and regulating a substantially greater number of PITX1 targets than is observed for ftz^{Q50K} and bcd, which is further supported by the fact that only $HOXD13^{Q317K}$ recognized the PITX1 binding site *in vitro*, indicating that the recognition of PITX1 sites is specific to $HOXD13^{Q317K}$. In addition, shared binding sites for PITX1 and $HOXD13^{Q317K}$ are enriched in the vicinity of the co-regulated genes, also indicating shared properties of $HOXD13^{Q317K}$ and PITX1.

4.3.2 Regulatory Consequences of Genome-Wide Binding Shifts are Mutation-Specific

An important aspect further influences the regulatory consequences of the $HOXD13^{Q317}$ mutations. The adoption of a novel set of target genes is only possible, if the new binding sites in the genome are meaningful in the regulatory context of the cell.

In the case of the $HOXD13^{Q317K}$ mutation, the cells in the embryonic limb bud are responsive to *Pitx1* expression, as indicated by *Pitx1* overexpression and knockout experiments in transgenic models (Logan and Tabin 1999; Szeto et al. 1999). In the patient, the altered binding of $HOXD13^{Q317K}$ is therefore likely to disturb the regulatory balance during limb bud development by regulating PITX1 targets in the *HOXD13* expression domain. Thus, a PITX1-like activity in the *HOXD13* expression domain can be expected to result in a major degree of misexpression. The hypothesis that normal HOXD13 function is partially replaced by PITX1 activity is supported by the effects of overexpressing $Hoxd13^{Q317K}$ and *PITX1* in chMM cultures and in infected chicken wing buds (see Section 0). The findings in chMM-cultures demonstrate that the mutant is able to bind and activate PITX1 target genes. The wing bud injections confirm that this effect is biologically relevant, as the mutant is able to induce PITX1-like morphological changes. The specificity of these effects is further corroborated by the investigation of the $HOXD13^{Q317R}$ mutant.

The shift in genomic binding introduced by the $HOXD13^{Q317R}$ mutation is not as drastic as seen for the $HOXD13^{Q317K}$ mutation and the limb bud injections show the effects of $HOXD13^{Q317R}$ to mostly overlap with $HOXD13^{wt}$. In addition to residual wildtype activity, the $HOXD13^{Q317R}$ recognition site does not correspond to known TF target sites and might therefore have only

minor effects. In comparison to previous analysis (Zhao et al. 2007), the new data highlights the potential insights and limitations of standard approaches. In light of the patient phenotype, the authors speculate about the possibility of a mixed LOF and GOF phenotype but cannot present any data supporting this. By using the chMM-ChIP-seq technique, a single experiment allowed not only to observe a genome-wide shift in binding, but also considerable residual wildtype HOXD13 activity, providing direct evidence for the hypothesis.

The underlying biochemistry of the HOXD13^{Q317K} and HOXD13^{Q317R} mutations (and of HOXD13^{I314L} (Caronia et al. 2003)) is the same. A change in amino acid composition in the recognition helix leads to an altered binding specificity of the TF. However, the resulting phenotypes differ drastically from one another. In comparison to a normal hand, the HOXD13^{I314L} mutation introduces the weakest change, whereas the HOXD13^{Q317R} phenotype is stronger and the HOXD13^{Q317K} phenotype deviates the most from a normal hand. Why is this? Caronia and colleagues (2003) demonstrated that the HOXD13^{I314L} recognizes a subset of HOXD13 motifs (a CTCGTAAAA sequence also typical for HOXA13) which in ChIP-seq would be expected to cause an even weaker shift in binding as compared to HOXD13^{Q317R} and HOXD13^{Q317K}. In context with the binding shifts observed for HOXD13^{Q317R} and HOXD13^{Q317K}, different degrees of remaining wildtype activity might partially explain the phenotype strength.
Thus, the results presented here clearly show that a missense mutation in a TF can act as gain-of-function mutation that shifts the binding profile of the relevant TF on a genome-wide scale. However, individual consequences of such a shift are dependent on the presence of compatible *cis*-regulatory elements/targets as well as the properties of protein domains involved in transactivation and protein-protein-interaction.

4.3.3 Despite low Overlap HOXD13^{Q317K} and PITX1 Show High Regulatory Similarity

Several factors contribute to the observation that, despite sharing only 1600 binding sites, HOXD13^{Q317K} and PITX1 induce highly similar chMM morphologies and global gene expression changes.
Importantly, the reported binding site overlap is likely to be an underestimation, since very stringent criteria were used for its calculation. However, even a more relaxed calculation of overlapping binding sites would lead to a maximum of approximately 5000 binding sites (15% of HOXD13^{Q317K} peaks). Likewise, despite the 25% overlap between HOXD13^{Q317R} and HOXD13^{wt} peaks, HOXD13^{Q317R} induces an almost indistinguishable phenotype in limb bud injections.

Moreover, similar to the results presented here, ChIP-seq experiments across model systems and organisms have reported thousands of binding sites for TFs across the genome (e.g. Ouyang et al. 2009; Junion et al. 2012; Wang et al. 2012b). Importantly, two recent studies suggest that a substantial amount of TF binding might be non-functional (Paris et al. 2013; Stefflova et al. 2013). These findings raise the question which fraction of the binding sites is functional and how to distinguish between non-functional and functional binding sites. Two promising approaches increase the likelihood of identifying functional binding sites. On the one hand, integration of multiple TF binding profiles in the same cells revealed that TFs tend to bind in clusters, generating highly occupied TF (HOT) regions (Ouyang et al. 2009; Gerstein et al. 2012; Junion et al. 2012; Kvon et al. 2012). When tested in reporter assays, these regions were shown to be much better indicators of functional *cis*-regulatory elements than, for example, evolutionary conserved regions (Kvon et al. 2012). On the other hand, identification of TF binding to functional *cis*-regulatory elements can be achieved by profiling a set of histone modifications that faithfully identify *cis*-regulatory elements in the sample of interest (Heintzman et al. 2007; Ernst and Kellis 2010).

Combining such methods with shared and differential binding of HOXD13^{Q317K} and PITX1 would help to identify those shared binding events that confer biological function. In principle, the chMM-ChIP-seq system is amenable to such kind of extension, which would provide further insights into the regulatory effects TF binding. However, additional experimental data would be required to further investigate the binding of and the regulation by wildtype and mutant HOXD13 proteins.

4.4 Analysis of the HOXD13^{R298Q} Mutation

4.4.1 The HOXD13^{R298Q} Phenotype in Relation to the Mild Loss-of-Function Found in ChIP-seq and Biochemical Approaches

The phenotype displayed by HOXD13^{R298Q} patients encompasses classical features of both SPD and BDA2. Whereas SPD phenotype is associated with several HOXD13 mutations, BDA2 symptoms have not yet been reported for HOXD13 mutations. A common effect of many HOXD13 mutations that cause SPD is the loss of functional HOXD13 protein through mRNA or protein degradation or loss of DNA-binding ability. The observed SPD in HOXD13^{R298Q} patients suggests that loss of HOXD13 function is likely one aspect of the pathomechanism. However, the unique BDA2 in HOXD13^{R298Q} patients strongly suggests that the protein function is altered in additional ways.

Interestingly, the results of all biological and biochemical assays performed in this study find the mutant protein to act as a hypomorph allele, since the loss in functionality compared to HOXD13^{wt} is only moderate. Contrary to the hypomorphic protein function, penetrance of HOXD13^{R298Q} in patients is stronger than seen for other loss-of-function mutations. All eight affected individuals display SPD symptoms and the additional BDA2 features were observed in five out of eight cases across three different families. Frameshift or nonsense mutations in HOXD13, however, are also associated with SPD, but mutation carriers are affected in as little as 30% of the cases (Goodman et al. 1998). Further complicating the genotype-phenotype relation, the HOXD13^{R298W} mutation modifies the same residue as in HOXD13^{R298Q}, but shows reduced penetrance as in haploinsufficient cases (Debeer et al. 2002). In the large HOXD13^{R298W} pedigree, only 3 in 17 mutation carriers show unilateral SPD and the remaining 14 individuals only exhibit a mild phenotype, a 5th finger clinodactyly. In this case, functional HOXD13 protein is possibly completely lost due to a strong reduction in DNA-binding ability (see Section 3.4.1 and 3.4.2).

In other words, whereas the HOXD13^{R298Q} mutation displays only a moderate loss-of-function in biological and biochemical experiments, it causes a highly penetrant phenotype, contrasting the effect of mutations that reduce more of the proteins “measurable” function. It is thus highly likely that the HOXD13^{R298Q} mutation exerts a deleterious, dominant-negative, effect on the wildtype allele in addition to the loss-of-function effects detected in experiments.

4.4.2 Possible Dominant-Negative Mechanisms of HOXD13^{R298Q} are Amenable to Experimental Investigation

By definition, a dominant-negative allele disrupts the activity of the wildtype allele in addition to losing its own functionality (Herskowitz 1987). Prominent examples of dominant-negative

effects are polyalanine tract expansions in HOXD13. When exceeding a certain length, the mutant HOXD13 proteins aggregate in the cytoplasm, are degraded, and also incorporate wildtype HOXD13 proteins into the aggregates, thereby reducing the amount of functional HOXD13 in the nucleus (Albrecht et al. 2004). In patients, the degree of the biochemical effect correlates with penetrance of the SPD phenotype (Goodman et al. 1997).

An experimental setup to test the dominant-negative potential of mutant HOXD13 was used by Caronia et al. (2003), who analyzed two homeodomain mutations. In that case, the authors showed that upon cotransfection of wildtype and mutant HOXD13, the mutant protein impaired the activation of a luciferase reporter construct by $HOXD13^{wt}$. Similar future experiments should validate whether the $HOXD13^{R298Q}$ mutant can act in a dominant-negative way over $HOXD13^{wt}$ (see below).

4.4.3 Molecular Mechanism Underlying Dominant-Negative Mutations

There are several biochemical possibilities how the mutant protein exerts its dominant-negative effect on a molecular level. Importantly, the type of biochemical mechanism that might underlie a dominant-negative effect can be experimentally addressed. The ChIP-seq analysis of $HOXD13^{R298Q}$ revealed some intriguing aspects that support discussion of possible pathomechanims.

AP1 and NFATc2 – Increased Cofactor Interaction?

Cofactor interaction is crucial for TF function; therefore, a possible effect of $HOXD13^{R298Q}$ might involve cofactor interaction. One possibility could be that reduced DNA-binding of $HOXD13^{R298Q}$ supports indirect binding events through cofactors. Given that those cofactors are DNA-binding TFs, increased interaction would be reflected in the sequence composition of the mutant's binding to sites when compared to the wildtype. Importantly, in such a situation the cofactor recognition sequences would be centrally enriched in those peaks.

The *de novo* detection of AP1 and NFATc2 motifs in predominantly $HOXD13^{R298Q}$-bound peaks suggested the possibility of indirect binding events mediated by AP1 or NFATc2 TFs. To further validate this hypothesis, Centrimo analysis was used to identify the distribution of AP1 and NFATc2 recognition sequences in $HOXD13^{R298Q}$ peaks (pp. 71-73). However, no central enrichment of the recognition sequences was found, contraindicating indirect DNA-binding of $HOXD13^{R298Q}$ via AP1 or NFATc2. In addition, interpretation of AP1 and NFATc2 motifs is problematic. AP1 TFs bind as dimers consisting of two subunits belonging to one of four protein families, cJun, cFos, ATF and JDP (reviewed in Hess et al. 2004). However, most of the possible heterodimeric complexes bind a similar motif. Moreover, AP1 dimers are extremely variable in their composition and, depending on composition, have been demonstrated to regulate gene expression in many biological processes. Thus, the presence of an AP1 binding

site is very challenging to interpret. In addition, chMM-ChIP-seq experiments in the group performed for the TFs MSX2 and RUNX2 also detect the AP1 motif in their peaks (Hein 2013). Since AP1 – Activating Protein 1 – has been shown to be a cofactor for many TFs (see Turpaev 2006 and references therein), one possible explanation might be, that AP1 motifs are present in a subset of peaks that is associated with (highly) transcribed genes, and therefore prone to be found in a broad range of ChIP-seq peaks.

For NFATc2, an indirect explanation for the *de novo* motif detection is even more likely. First, the GGAAA motif is only a half-site of the normally dimeric binding sites (Jolma et al. 2013). Second, the motif was mainly found in weakly bound HOXD13^{R298Q} binding sites. Finally, the sequence composition is unspecific, so that the detection threshold for FIMO analysis had to be lowered. However, a functional role for NFATc2 cannot fully be excluded, since the gene is strongly expressed in developing cartilage and can repress chondrogenesis (Ranger et al. 2000). Therefore, the NFATc2 motif possibly represents an unspecific signal that becomes visible in weakly enriched binding sites.

Taken together, a closer analysis of both motifs detected in predominantly HOXD13^{R298Q} bound peaks does not indicate altered (stronger) interaction of the mutant with AP1 or NFATc2 proteins.

HOXA13 – Decreased Cofactor Interaction?

In contrast to AP1 and NFATc2, HOXA13 binding sites were reduced in HOXD13^{R298Q} peaks. The large fraction of HOXD13^{wt} peaks containing a HOXA13 binding site was a surprising and characteristic outcome of the HOXD13^{wt} ChIP-seq (discussed below), and was not observed for HOXD13^{R298Q} ChIP-seq. The fraction of HOXD13^{R298Q} peaks containing a HOXA13 motif was lower in general and specifically the number of peaks containing motifs for both – HOXD13 and HOXA13 – was strongly reduced. A possible explanation could be that the R298Q mutation hinders putative protein-protein interaction between HOXD13 and HOXA13 proteins. Homeodomains are known to form dimers, however, the interaction between homeodomains is usually mediated via the N-terminal arm or non-homeodomain sequences of the respective proteins (reviewed in Mann et al. 2009). Additionally, detailed analysis of dimeric homeodomain-complexes demonstrated that the recognition sequences of dimeric homeodomain complexes differ decisively from those of the monomers (Joshi et al. 2007; Slattery et al. 2011). Assuming that HOXD13-HOXA13-both peaks are more likely to represent functional biding sites, the reduction of those peaks could be due to an indirect effect, given that HOXD13^{R298Q} fails to distinguish between functional and non-functional biding sites.

However, some literature reports support a possible direct effect of the missense mutation on the observed reduction in HOXD13-HOXA13 both peaks. Saadi and colleagues (2003) described a dominant-negative acting mutation in the homeodomain TF *PITX2* in patients with Axenfeld-

Rieger syndrome. In carefully designed experiments, the authors demonstrated that the mutant protein hindered dimeric PITX2-complexes, composed of wildtype and mutant PITX2 proteins, from adequately binding to and activating reporter constructs. In another line of evidence, recent experiments demonstrated HOXA13 homeodomains binding to DNA as a dimer. Interestingly, interaction between both homeodomains was mediated via the second helix that harbors the mutated R298 (Zhang et al. 2011). Although the R298 is not involved in dimer formation and dimeric DNA-binding of HOXD13 has not been shown, a dominant-negative effect of $HOXD13^{R298Q}$ on this binding might be a possible pathomechanism and could be addressed experimentally using coimmunoprecipitation, EMSAs or similar biochemical experiments. However, contradictory evidence for this line of argumentation is provided by the $HOXD13^{R298W}$ mutation that does not act dominant-negatively. In this mutation, the identical residue is changed to a bulky tryptophan, which is why the R298W substitution might disturb protein-protein interaction even more than the R298Q substitution and should thereby provoke a similar dominant-negative effect.

Collectively, the sequence composition of $HOXD13^{R298Q}$ peaks give some indications towards altered cofactor interaction; however, further studies are needed to assess whether the detected differences are in fact due to altered protein-protein interaction of the mutant or an indirect effect as a consequence failure to distinguish between functional and non-functional binding sites.

Dominant-Negative Effects on $HOXD13^{wt}$ via Competitive Binding

The sum of experimental evidence strongly suggests that $HOXD13^{R298Q}$ acts as a hypomorph allele in terms of DNA-binding and regulation, whereas it likely has a dominant-negative effect in the patients. $HOXD13^{R298W}$, in contrast, shows no such discrepancy between experimental and medical evidence and is highly likely acting as an amorphic allele, retaining no regulatory function (Debeer et al. 2002). The question arises how a hypomorphic allele exerts a stronger phenotype than an amorphic one, and whether the mechanism could be independent of cofactor interactions?

To answer the question, the mutated protein's biochemical function, in this case the TF HOXD13, must be taken into account. Aside from isolated reports (Zhang et al. 2011), overwhelming evidence suggests that HOX-proteins do not bind DNA as a homodimer (Gehring et al. 1994; Berger et al. 2008; Jolma et al. 2013). Therefore, mechanisms distinct from protein-protein interaction must be evaluated. Veitia (2007) described the seemingly paradoxical effect of hypomorph mutations based on the assumption of cooperative binding of TFs. Cooperative or synergistic binding is independent of dimer formation and thought to be a common mechanism through which TFs activate their target genes. This is strongly supported

by clustered occurrence of TF binding sites in regulatory elements that remain functional independent of orientation or distance of the binding sites to one another (Arnosti et al. 1996; Swanson et al. 2010). In a simulation of cooperative activation, a hypomorph mutation caused a much stronger reduction of transcriptional output than the haploinsufficient situation. The dominant-negatively acting missense mutations in PITX2 supported this simulation and also a dominant-negative mutation in KLF1 has been interpreted to act through such a mechanism (Saadi et al. 2003; Siatecka et al. 2010).

Another aspect is that not only sequence specificity, but also residency at the bound DNA sequence is of importance to TF function. In an elegant ChIP-seq experiment Lickwar and colleagues (2012) recently demonstrated differences between TF binding sites that were very briefly bound and sites that were bound for longer time period. Strikingly, the sites that were bound for a long time were much more likely to be associated with genes that were transcriptionally regulated by the TF than the briefly bound sites.

In fact, the large reduction of HOXD13-HOXA13 both peaks and the reduced DNA-binding affinity of HOXD13^{R298Q} suggest that (1) HOXD13^{R298Q} might specifically lose binding sites that require cooperative regulation and (2) the reduced binding strength itself hinders the regulatory potential of the mutant protein.

Thus, a dominant-negative effect of the hypomorph HOXD13^{R298Q} may result from a competition between the mutant and the wildtype HOXD13 at each individual binding site. If the HOXD13^{R298Q} binds to a site, its reduced affinity would lead to a premature dissociation from the DNA preventing efficient regulation of target genes. Importantly, at the same time the binding event would prevent the wildtype protein from binding to DNA. In combination, the hypomorph HOXD13^{R298Q} would bind long enough to disturb HOXD13^{wt} function but not long enough to adequately regulate gene expression. Analogously, the DNA-binding of the HOXD13^{R298W} would be so severely hampered, that the mutant does compete with HOXD13^{wt} for DNA-binding, thereby acting as a haploinsufficient allele.

However, it needs to be noted that the dominant-negative effects discussed above do not explain the BDA2 phenotype observed in the patients, which might result from additional effects of the HOXD13^{R298Q} protein that would be independent of the putative dominant-negative mechanism discussed here.

Taken together, the analysis and interpretation of the HOXD13^{R298Q} mutation turns out to be more complicated than the effect introduced by altered sequence specificity in HOXD13^{Q317} mutations. However, preliminary results point to highly interesting insights regarding HOXD13 function and warrant further investigation.

4.4.4 Outlook

Analysis of HOXD13^{R298Q} binding sites especially in relation to the strong phenotype suggests a dominant-negative effect of the mutant protein. However, the experiments performed so far have not proven this hypothesis and future experiments need to validate the initial findings.

This could be achieved by performing experiments similar to the luciferase experiments performed by Caronia et al. (2003). The chMM system provides a convenient experimental setup to test a possible dominant negative effect of HOXD13 mutations. Double infection of chMM cultures using RCASBP(A) and RCASBP(B) vectors leads to expression of both wildtype and mutant HOXD13 in the same cell. In an experimental setup, infection with RCASBP(B)-virus expressing a HOXD13^{wt} will reduce chondrogenic differentiation of the chMM-culture to a measurable extent. Modulation of this effect could then be measured by coinfection of RCASBP(B)-HOXD13^{wt} and RCASBP(A)-HOXD13^{R298Q} viruses. If HOXD13^{R298Q} acts dominant-negatively over HOXD13^{wt}, the inhibition of chondrogenic differentiation by HOXD13^{wt} would be reduced in this assay. Three possible control experiments should be performed to validate the result: coinfection with additional HOXD13^{wt}, HOXD13^{+7Ala}, or HOXD13^{R298W} RCASBP(A)-viruses. HOXD13^{wt} coinfection would further reduce chondrogenic differentiation, HOXD13^{+7Ala} would also act dominant-negative, and HOXD13^{R298W} should not modify the effect of HOXD13^{wt}, since it is thought to act as a loss-of-function allele with no dominant-negative effect. A further advantage of using the chMM-system for this approach is that the measured effect on chondrogenic differentiation is systemic and independent of the regulatory requirements at individual reporter constructs used in luciferase assays.

Investigation of HOXD13-HOXA13 both peaks showed an intriguing difference between HOXD13^{R298Q} and HOXD13^{wt}. Further experiments using EMSAs should investigate whether (1) HOXD13^{wt} can bind DNA as a dimer, (2) the orientation of binding sites towards one another is influencing this binding and (3) whether HOXD13^{R298Q} and/ HOXD13^{R298W} variants show a different behavior in any of these experiments. The outcome of such experiments should allow important insights into the pathomechanism for both HOXD13^{R298} mutations.

4.5 Interpretation of Genome-wide TF Binding

Despite intensive studies, the molecular mechanism of TF function remains rather mysterious. The genome-wide binding data with thousands of binding sites presented in this study is in line with the findings of many other groups reporting on the binding of various TFs (e.g. Ouyang et al. 2009; Junion et al. 2012; Wang et al. 2012b). It remains difficult to assign ChIP-seq peaks to target genes as well as functionality to individual ChIP-seq peaks (reviewed in MacQuarrie et al. 2011; Spitz and Furlong 2012). As in our results, especially developmental TFs have been

shown to bind to numerous sites throughout the genome. Further complicating the analysis, many peaks are found in intergenic regions that cannot be easily attributed to the regulation of a specific target (Cao et al. 2010). Therefore the absence of a nearby peak does not exclude direct regulation of a certain gene through a TF. However, neither does the presence of ChIP-seq peaks in promoters of differentially expressed genes prove direct regulation. Thus, although for example the enrichment of HOXD13^{Q317K}-PITX1 shared peaks near co-regulated genes links binding to regulation, the list of candidate genes generated in this manner will neither be complete nor will all the genes be direct targets.

In recent years, various approaches have been used to increase the probability of identifying functional binding sites by adding information to an initial ChIP-seq data set (see also Section 4.3.3), such as restricting TF binding sites to those located in evolutionary conserved elements (Visel et al. 2008), combining binding data for multiple, functionally related TFs (Junion et al. 2012), or generating a chromatin profile of the ChIP sample (Heintzman et al. 2007; Ernst and Kellis 2010 and 2013). The results gained by these approaches show, that additional information is needed to interpret the genomic TF binding sites.

In that respect, the design of the chMM-ChIP-seq experiments performed here limits the interpretation of functional TF binding in chMM cultures, because additional filters to interpret the ChIP-seq peaks are missing. However, the chMM-ChIP-seq system established in this study enables the design of experiments that could add missing datasets to address a wide range of biological questions.

4.6 General Implications for HOXD13 Binding and Function

This study provides the first report of a HOXD13 ChIP-seq dataset and finds HOXD13 to bind to thousands of sites throughout the genome and recognizes the same DNA-sequence as the homeodomain *in vitro* (CCAATAAAA). Strikingly, approximately 22% of HOXD13 peaks contain the reported HOXA13 recognition sequence (CTCGTAAAA) (Berger et al. 2008; Zhang et al. 2011; Jolma et al. 2013). These peaks either contain only a HOXA13 site (HOXA13-only) or a HOXA13 site in addition to a HOXD13 binding site (HOXD13-HOXA13 both). This finding confirms previous observations by several groups that found the HOXD13 homeodomain to bind to the HOXA13 recognition sequence and vice versa in *in vitro* assays (Caronia et al. 2003; Berger et al. 2008; Zhang et al. 2011; Jolma et al. 2013). An additional key observation is that one in five HOXD13 motifs found in HOXD13 peaks contained an additional HOXA13 binding site within 300 bp, which is also reflected in the characteristic distribution of the best HOXD13 sequence found in Centrimo analysis. Notably, the HOXA13 motif used for the analysis is highly similar to the motif that HOXD12, HOXD11 and HOXA11 bind *in vitro* (Berger et al. 2008; Jolma et al. 2013). The similarity of the motifs complicates the conclusions concerning cobinding of any two HOX proteins. Therefore, the results do not allow

to conclude that HOXD13 specifically binds with HOXA13 throughout the genome, but rather suggest a general cooperative regulation by all posterior HOX proteins.

This observation is in line with the phenotype of *Hoxd11-13* triple knockout mice, which exhibit a syndactyly that strongly resembles SPD (Zakany and Duboule 1996). Similarly, HOXD13 and HOXA13 redundantly activate the EphA7 promoter in reporter constructs and inhibit chondrogenic differentiation in chMM, respectively (Salsi and Zappavigna 2006; Kuss et al. 2009). Detailed analysis of the HOXD13-HOXA13 motif pairs provided little evidence for dimeric binding of two HOX-proteins but rather suggest a cooperative binding. However, in a minority of cases, the HOXD13-HOXA13 recognition sites seem to exhibit a preferred orientation, possibly indicating steric limitations to the binding of two HOX proteins to directly adjacent binding sites.

The analyses performed over the course of this study did not focus on the regulatory effect of HOXD13 but rather on its biochemical binding properties. Salsi et al. (2008) performed a ChIP-chip study of endogenous and transiently expressed *Hoxd13* in murine limb buds and human osteogenic cell lines, respectively, and demonstrated that HOXD13 binds and regulates genes involved in limb development and/or cartilage/bone formation. The chMM-ChIP-seq system offers an excellent opportunity to study skeletal development and future analyses using chMM-ChIP-seq should provide insights into the mechanisms and regulatory pathways through which HOXD13 influences chondrogenesis.

Collectively, analysis of the HOXD13 binding profile strongly suggests that posterior HOX proteins bind cooperatively throughout the genome and confirms biological evidence of posterior HOX proteins' redundancy.

4.7 The End

Using chMM-ChIP-seq, distinct pathomechanisms underlying two HOXD13 mutations were uncovered. First, the HOXD13^{Q317K} mutation was found to specifically change the sequence specificity, thereby inducing a genome-wide shift in binding. Interestingly, the regulatory effects of such a genomic shift in binding are connected to the presence or absence of other TF binding sites that make this shift meaningful.

Second, the HOXD13^{R298Q} mutation reduced the DNA-binding affinity of the mutant protein, possibly causing a loss of functional binding sites. A key observation regarding this loss is that the pathogenic potential of the mutant HOXD13 does not seem to be directly connected to the binding strength. However, the experimental data presented here are incomplete and warrant further investigation.

Finally, a close analysis of HOXD13 binding sites revealed a high fraction of binding sites with several HOX-like recognition sequences, which points to a molecular basis for the functional redundancy of this TF family.

Table 5.1 Data underlying the quality control and reproducibility tests for ChIP-seq experiments

Identifier	# seq reads	QC filtered reads	# uniq map reads	# non-red, reads	Self Cons IDR 0,01 (Top120k)	Bio-Rep IDR 0,01	Pooled PseudoRep IDR 0,0025	PBC	NSC	RSC
HOXD13_A(2010-08-31-1)_Rep1	61,196,225	37,203,825	30,595,807	23,207,443	-	-	-	0.76	1,118	2.68
HOXD13_A(2010-08-31-2)_Rep1	62,819,402	38,179,703	31,010,014	23,483,060	-	-	-	0.76	1.117	2.66
HOXD13_A_Rep1(TR)	124,015,627	75,383,528	61,515,821	46,690,503	16,385	22,167	34,267	0.76	1.118	2.65
HOXD13wt_B(2011-05-18-1)_Rep2	58,145,834	39,334,430	35,139,109	31,067,003	29,555	22,167	34,267	0.88	1.087	3.49
Q317K_A(2011-05-18-3)_Rep1	57,715,872	40,904,973	36,212,424	29,738,115	31,805	16,167	25,036	0.82	1.124	2.28
Q317K_B(2010-06-04-8)_Rep2	31,319,062	25,375,204	21,362,610	20,096,638	32,420	16,167	25,036	0.94	1.071	1.66
Q317R_A(2011-05-18-4)_Rep1	56,321,020	39,614,344	35,441,509	30,532,308	28,181	3,404	21,396	0.86	1.107	2.15
Q317R_B(2012-12-05-7)_Rep2	49,919,155	39,333,594	30,865,446	24,779,434	18,465	3,404	21,396	0.80	1.042	1.53
PITX1_A(2011-04-15-3)_Rep1	52,858,471	36,677,369	31,135,297	28,465,840	40,632	39,123	40,881	0.91	1.203	1.43
PITX1_B(2011-05-18-5)_Rep2	40,118,036	33,918,205	23,969,984	19,418,256	-	-	-	0.81	1.260	1.05
PITX1_B(2011-12-01-5)_Rep2	36,249,445	31,594,322	22,417,436	18,403,140	-	-	-	0.82	1.262	1.06
PITX1_B_Rep2(TR)	76,367,481	65,512,527	46,387,420	37,821,396	36,115	39,123	40,881	0.82	1.261	1.05

Identifier	# seq reads	QC filtered reads	# uniq map reads	# non-red, reads
HOXD13wt_A_INPUT(2012-01-10-5)	55,158,595	38,795,156	32,243,813	31,409,557
HOXD13wt_B_INPUT(2012-01-10-4)	56214,309	40,353,549	33,877,719	11,800,101
Q317K_A_INPUT(2012-01-10-7)	50,070,245	38,967,549	31,292,292	29,609,156
Q317K_B_INPUT(2012-01-10-8)	46,591,856	36,492,722	26,903,729	13,700,381
Q317R_A_INPUT(2012-11-06-5)	54,806,846	40,871,471	31,756,019	30,904,530
Q317R_B_INPUT(2012-10-30-4)	42,594,880	33,337,274	22,415,339	19,103,707
cPITX1_A_INPUT(2012-01-10-3)	57,801,527	38,699,029	32,080,090	30,018,125
cPITX1_B_INPUT(2012-01-10-2)	52,832,421	38,179,037	31,429,645	24,874,733

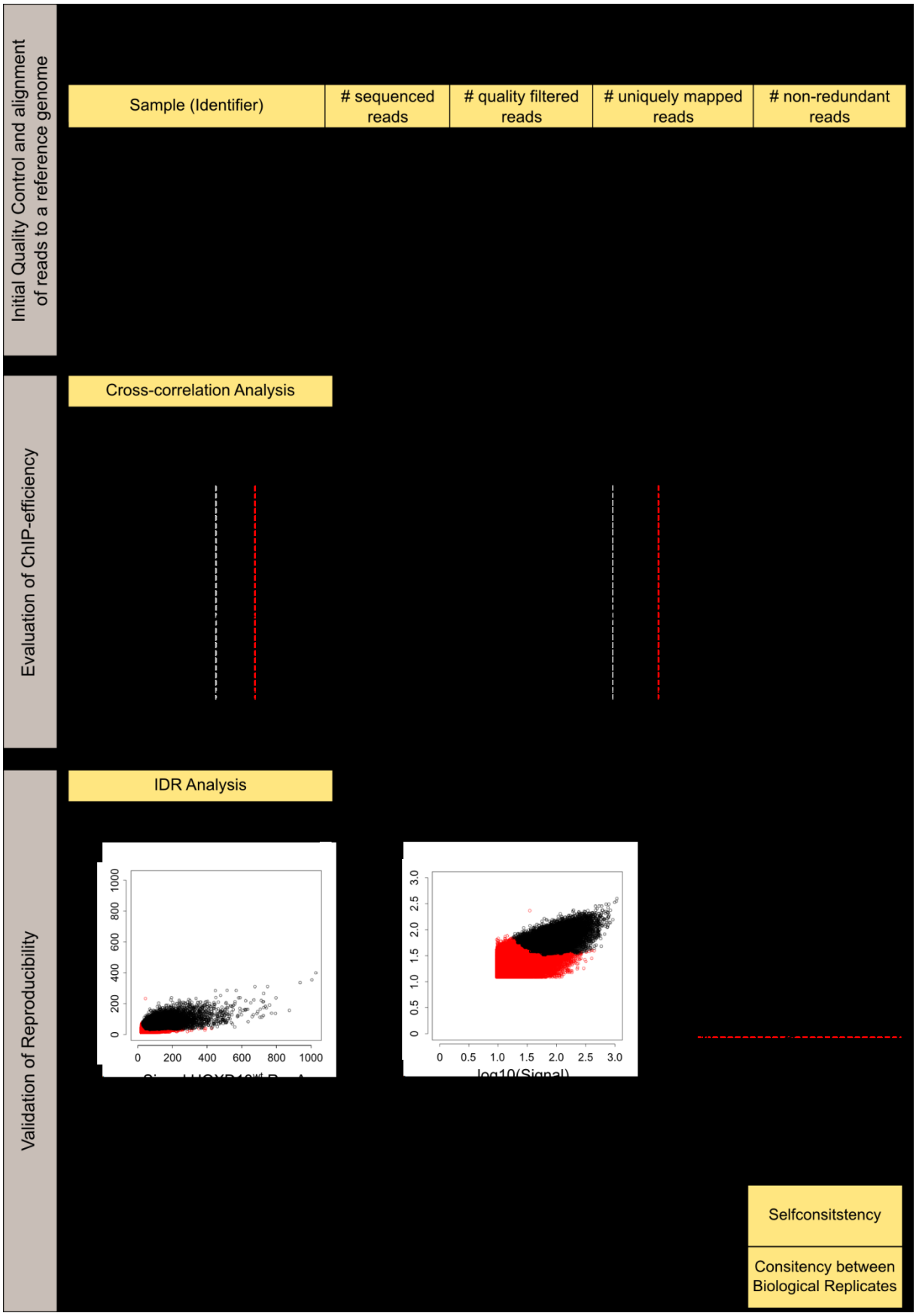


Figure 5.1 Data sheet for HOXD13^{wt} ChIP-seq.

Quality-filtering and read-mapping, cross-correlation analysis and IDR analysis for HOXD13^{wt} replicates performed as described in Section 2.3.6.

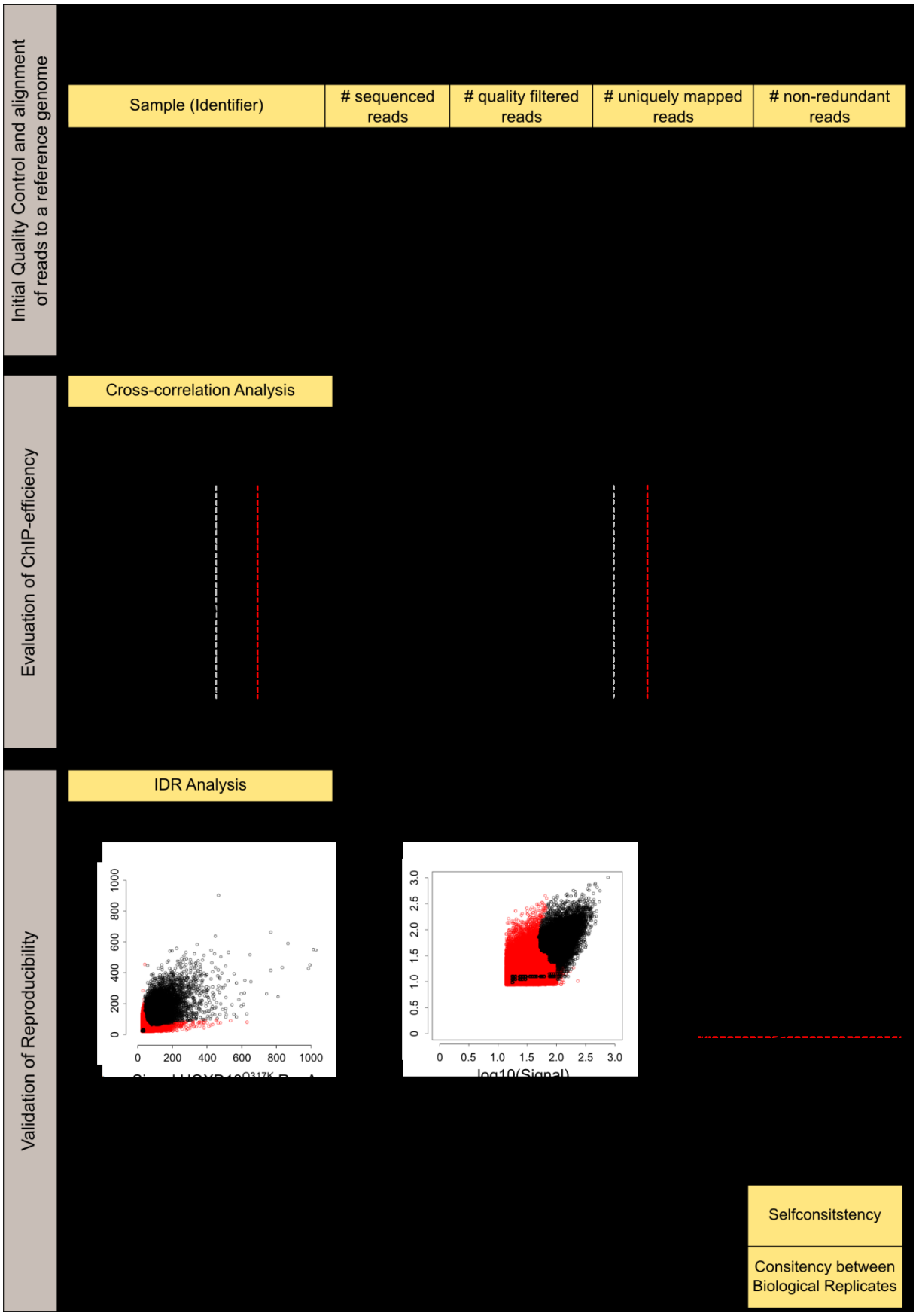


Figure 5.2 Data sheet for HOXD13^{Q317K} ChIP-seq.

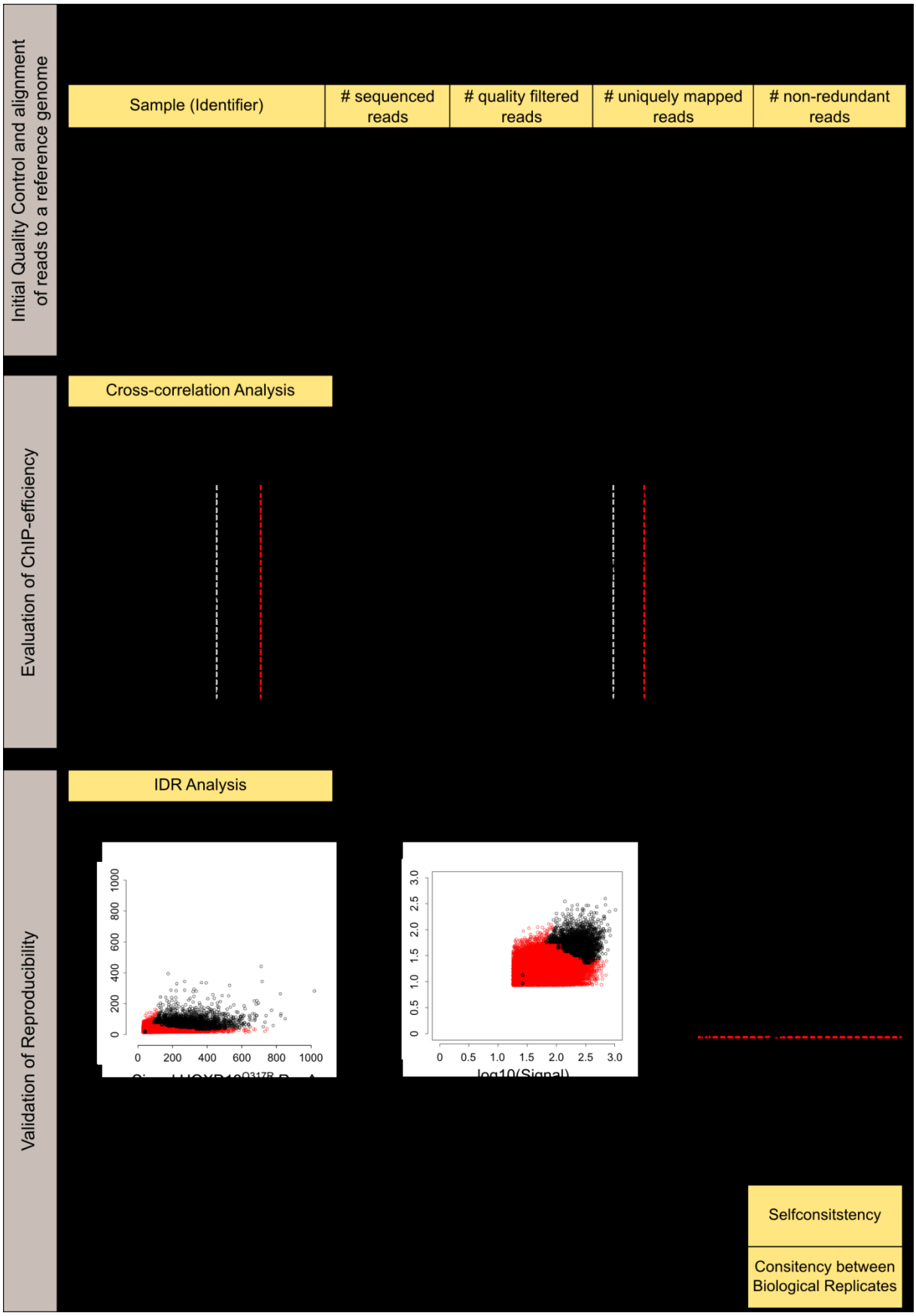


Figure 5.3 Data sheet for HOXD13^{Q317R} ChIP-seq.

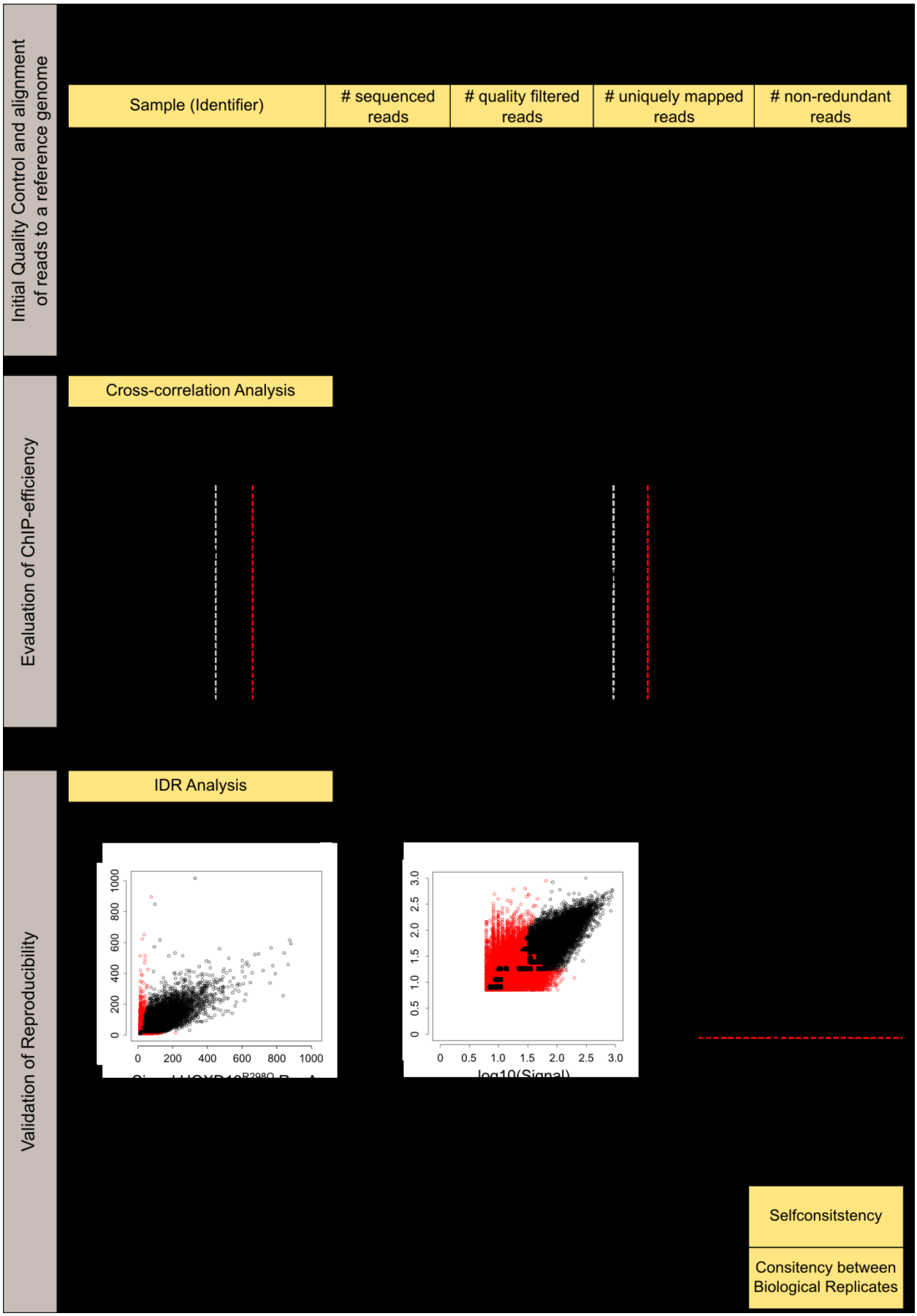


Figure 5.4 Data sheet for HOXD13^{R298Q} ChIP-seq.

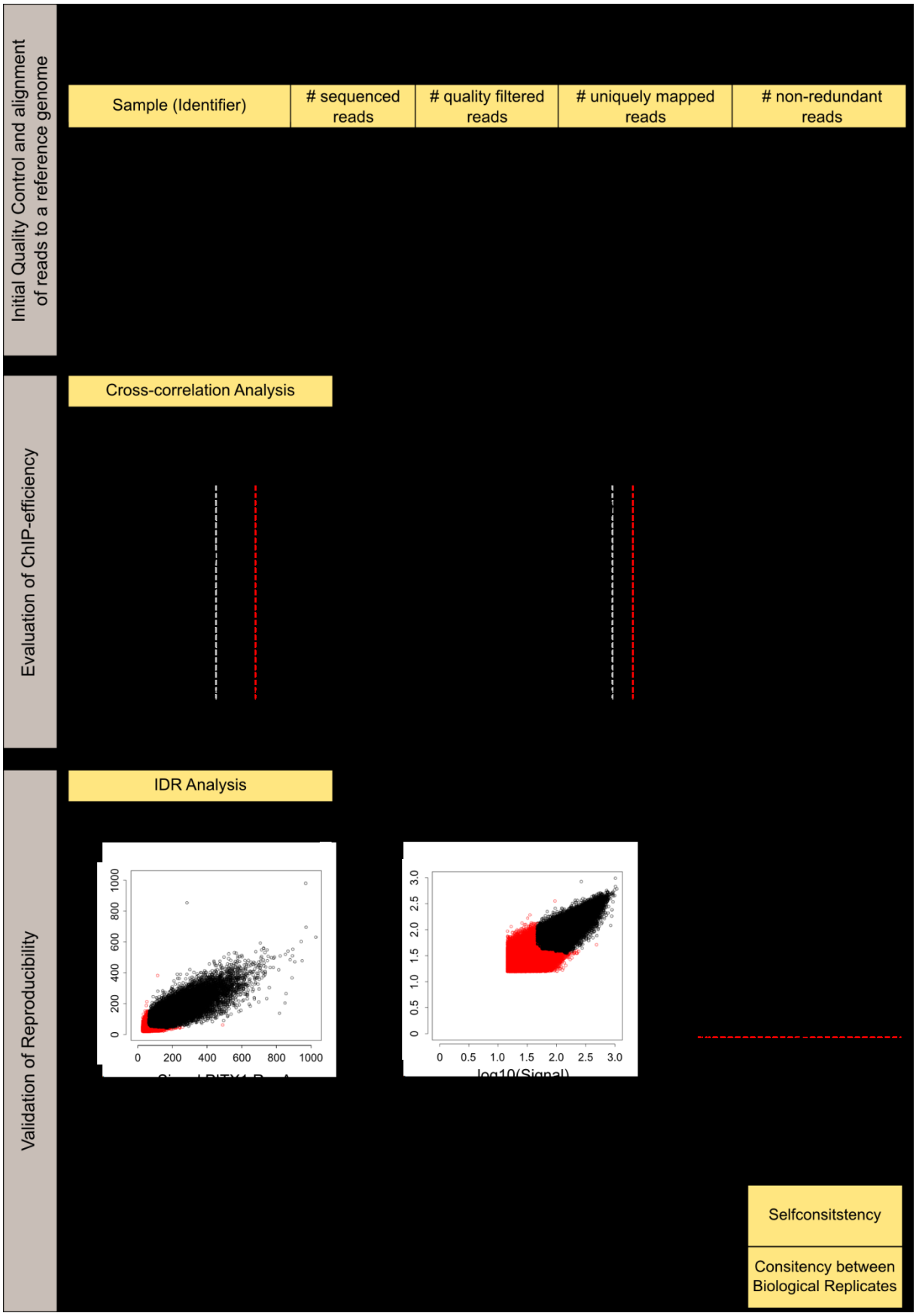


Figure 5.5 Data sheet for PITX1 ChIP-seq.

6 References

Akam M. 1989. Hox and HOM: Homologous gene clusters in insects and vertebrates. *Cell* **57**(3): 347-349.

Albrecht AN, Kornak U, Böddrich A, Süring K, Robinson PN, Stiege AC, Lurz R, Stricker S, Wanker EE, Mundlos S. 2004. A molecular pathogenesis for transcription factor associated poly-alanine tract expansions. *Human Molecular Genetics* **13**(20): 2351-2359.

Arnosti DN, Barolo S, Levine M, Small S. 1996. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**(1): 205-214.

Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**(12): 1653-1659.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**(suppl 2): W202-W208.

Bamshad M, Le T, Watkins WS, Dixon ME, Kramer BE, Roeder AD, Carey JC, Root S, Schinzel A, Van Maldergem L et al. 1999. The spectrum of mutations in TBX3: Genotype/Phenotype relationship in ulnar-mammary syndrome. *The American Journal of Human Genetics* **64**(6): 1550-1562.

Bamshad M, Lin RC, Law DJ, Watkins WS, Krakowiak PA, Moore ME, Franceschini P, Lala R, Holmes LB, Gebuhr TC. 1997. Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat Genet* **16**(3): 311-315.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET et al. 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**(7): 1266-1276.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology / edited by Frederick M Ausubel [et al]* **Chapter 19**.

Boyadjiev SA, Jabs EW. 2000. Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clinical Genetics* **57**(4): 253-266.

Brison N, Debeer P, Fantini S, Oley C, Zappavigna V, Luyten FP, Tylzanowski P. 2012a. An N-terminal G11A mutation in HOXD13 causes synpolydactyly and interferes with Gli3R function during limb pre-patterning. *Human Molecular Genetics* **21**(11): 2464-2475.

Brison N, Debeer P, Tylzanowski P. 2013. Joining the fingers- A HOXD13 story. *Dev Dyn*: n/a-n/a.

Brison N, Tylzanowski P, Debeer P. 2012b. Limb skeletal malformations - What the HOX is going on? *European Journal of Medical Genetics* **55**(1): 1-7.

Bulyk ML, Huang X, Choo Y, Church GM. 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* **98**(13): 7158-7163.

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL et al. 2010. Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. *Developmental Cell* **18**(4): 662-674.

Caronia G, Goodman FR, McKeown CME, Scambler PJ, Zappavigna V. 2003. An I47L substitution in the HOXD13 homeodomain causes a novel human limb malformation by producing a selective loss of function. *Development* **130**(8): 1701-1712.

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**(6540): 479-485.

Chen Y-C, Rajagopala SV, Stellberger T, Uetz P. 2010. Exhaustive benchmarking of the yeast two-hybrid system. *Nat Meth* **7**(9): 667-668.

Chlon Timothy M, Doré Louis C, Crispino John D. 2012. Cofactor-Mediated Restriction of GATA-1 Chromatin Occupancy Coordinates Lineage-Specific Gene Expression. *Molecular Cell* **47**(4): 608-621.

D'Haeseleer P. 2006. What are DNA sequence motifs? *Nat Biotech* **24**(4): 423-425.

Davidson EH. 2006. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Academic Press.

De Folter S, Immink RGH. 2011. Yeast protein-protein interaction assays and screens. Vol 754, pp. 145-165.

Deane CM, Salwiński Ł, Xenarios I, Eisenberg D. 2002. Protein Interactions: Two Methods for Assessment of the Reliability of High Throughput Observations. *Molecular & Cellular Proteomics* **1**(5): 349-356.

Debeer P, Bacchelli C, Scambler PJ, De Smet L, Fryns JP, Goodman FR. 2002. Severe digital abnormalities in a patient heterozygous for both a novel missense mutation in HOXD13 and a polyalanine tract expansion in HOXA13. *J Med Genet* **39**(11): 852-856.

Dollé P, Dierich A, LeMeur M, Schimmang T, Schuhbaur B, Chambon P, Duboule D. 1993. Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* **75**(3): 431-441.

Dolle P, Izpisua-Belmonte J-C, Falkenstein H, Renucci A, Duboule D. 1989. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature* **342**(6251): 767-772.

Dorman CM, Johnson SE. 1999. Activated Raf inhibits avian myogenesis through a MAPK-dependent mechanism. *Oncogene* **18**(37): 5167-5176.

Duboc V, Logan MPO. 2011. Regulation of limb bud initiation and limb-type morphology. *Dev Dyn* **240**(5): 1017-1027.

Dudley AT, Ros MA, Tabin CJ. 2002. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **418**(6897): 539-544.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Computational Biology* **28**(8): 817-838.

-. 2012. ChromHMM: automating chromatin-state discovery and characterization.

Ernst J, Kellis M. 2013. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Research*.

Etchevers H. 2011. Primary culture of chick, mouse or human neural crest cells. *Nat Protocols* **6**(10): 1568-1577.

Fantini S, Vaccari G, Brison N, Debeer P, Tylzanowski P, Zappavigna V. 2009. A G220V substitution within the N-terminal transcription regulating domain of HOXD13 causes a variant synpolydactyly phenotype. *Human Molecular Genetics* **18**(5): 847-860.

Fraenkel E, Rould MA, Chambers KA, Pabo CO. 1998. Engrailed homeodomain-DNA complex at 2.2 å resolution: a detailed view of the interface and comparison with other engrailed structures. *Journal of molecular biology* **284**(2): 351-361.

Furniss D, Kan S-h, Taylor IB, Johnson D, Critchley PS, Giele HP, Wilkie AOM. 2009. Genetic screening of 202 individuals with congenital limb malformations and requiring reconstructive surgery. *J Med Genet* **46**(11): 730-735.

Garcia-Barceló M-M, Wong KK-y, Lui VC-h, Yuan Z-w, So M-t, Ngan ES-w, Miao X-p, Chung PH-y, Khong P-l, Tam PK-h. 2008. Identification of a HOXD13mutation in a VACTERL patient. *American journal of medical genetics Part A* **146A**(24): 3181-3185.

Gehring WJ, Kloter U, Suga H. 2009. Chapter 2 Evolution of the Hox Gene Complex from an Evolutionary Ground State. In *Current Topics in Developmental Biology*, Vol Volume 88 (ed. P Olivier), pp. 35-61. Academic Press.

Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K. 1994. Homeodomain-DNA recognition. *Cell* **78**(2): 211-223.

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R et al. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**(7414): 91-100.

Goecks J, Nekrutenko A, Taylor J, The Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**(8): R86.

Gong K-Q, Yallowitz AR, Sun H, Dressler GR, Wellik DM. 2007. A Hox-Eya-Pax Complex Regulates Early Kidney Developmental Gene Expression. *Molecular and Cellular Biology* **27**(21): 7661-7668.

Goodman F, Giovannucci-Uzielli ML, Hall C, Reardon W, Winter R, Scambler P. 1998. Deletions in HOXD13 segregate with an identical, novel foot malformation in two unrelated families. *The American Journal of Human Genetics* **63**(4): 992-1000.

Goodman FR. 2002. Limb malformations and the humanHOX genes. *American journal of medical genetics* **112**(3): 256-265.

Goodman FR, Mundlos S, Muragaki Y, Donnai D, Giovannucci-Uzielli ML, Lapi E, Majewski F, McGaughran J, McKeown C, Reardon W et al. 1997. Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proceedings of the National Academy of Sciences* **94**(14): 7458-7463.

Goodman FR, Scambler PJ. 2001. Human HOX gene mutations. *Clin Genet* **59**(1): 1-11.

Graham A, Papalopulu N, Krumlauf R. 1989. The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* **57**(3): 367-378.

Haugen SP, Ross W, Gourse RL. 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Micro* **6**(7): 507-519.

He A, Kong SW, Ma Q, Pu WT. 2011. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proceedings of the National Academy of Sciences* **108**(14): 5632-5637.

Hein H. 2013. Functional analysis of transcription factors in chondrogenic and osteogenic differentiation using ChIP-seq. In *FU Berlin*.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3): 311-318.

Herskowitz I. 1987. Functional inactivation of genes by dominant negative mutations. *Nature* **329**(6136): 219-222.

Hess J, Angel P, Schorpp-Kistner M. 2004. AP-1 subunits: quarrel and harmony among siblings. *Journal of Cell Science* **117**(25): 5965-5973.

Hill D, Hope I, Macke J, Struhl K. 1986. Saturation mutagenesis of the yeast his3 regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234**(4775): 451-457.

Holmqvist P-H, Boija A, Philip P, Crona F, Stenberg P, Mannervik M. 2012. Preferential Genome Targeting of the CBP Co-Activator by Rel and Smad Proteins in Early <italic>Drosophila melanogaster</italic> Embryos. *PLoS Genet* **8**(6): e1002769.

Huang DW, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols* **4**(1): 44-57.

-. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**(1): 1-13.

Infante CR, Park S, Mihala A, Kingsley DM, Menke DB. 2012. Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. *Developmental Biology*: 1-11.

Jamsheer A, Sowińska A, Kaczmarek L, Latos-Bieleńska A. 2012. Isolated brachydactyly type E caused by a HOXD13 nonsense mutation: a case report. *BMC Medical Genetics* **13**(1): 4.

Johnson D, Kan S-H, Oldridge M, Trembath RC, Roche P, Esnouf RM, Giele H, Wilkie AOM. 2003. Missense mutations in the homeodomain of HOXD13 are associated with brachydactyly types D and E. *The American Journal of Human Genetics* **72**(4): 984-997.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **316**(5830): 1497-1502.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta Kazuhiro R, Rastas P, Morgunova E, Enge M, Taipale M, Wei G et al. 2013. DNA-Binding Specificities of Human Transcription Factors. *Cell* **152**(1–2): 327-339.

Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS. 2007. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* **131**(3): 530-543.

Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EEM. 2012. A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. *Cell* **148**(3): 473-486.

Kadonaga JT. 2004. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* **116**(2): 247-257.

Kan S-H, Johnson D, Giele H, Wilkie AOM. 2003. An acceptor splice site mutation inHOXD13 results in variable hand, but consistent foot malformations. **121A**(1): 69-74.

Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. **26**(12): 1351-1359.

Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**(suppl 2): W43-W46.

Kmita M, Tarchini B, Zakany J, Logan M, Tabin CJ, Duboule D. 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature* **435**(7045): 1113-1116.

Kohlhase J, Wischermann A, Reichenbach H, Froster U, Engel W. 1998. Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nat Genet* **18**(1): 81-83.

Komuro I, Izumo S. 1993. Csx: a murine homeobox-containing gene specifically expressed in the developing heart. *Proceedings of the National Academy of Sciences* **90**(17): 8145-8149.

Kundaje A. 2012. https://sites.google.com/site/anshulkundaje/projects/idr. Vol accessed 2013-03-01.

Kurban M, Wajid M, Petukhova L, Shimomura Y, Christiano AM. 2011. A nonsense mutation in the HOXD13 gene underlies synpolydactyly with incomplete penetrance. *J Hum Genet* **56**(10): 701-706.

Kuss P, Villavicencio-Lorini P, Witte F, Klose J, Albrecht AN, Seemann P, Hecht J, Mundlos S. 2009. Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *Journal of Clinical Investigation* **119**(1): 146-156.

Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ, Stark A. 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature.

Lanctot C, Moreau A, Chamberland M, Tremblay ML, Drouin J. 1999. Hindlimb patterning and mandible development require the Ptx1 gene. *Development* **126**(9): 1805-1810.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* **22**(9): 1813-1831.

Lee TI, Johnstone SE, Young RA. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. **1**(2): 729-748.

Lee TI, Young RA. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**: 77-137.

Lelli KM, Slattery M, Mann RS. 2012. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. *Annu Rev Genet* **46**(1): 43-68.

Lenhard B, Sandelin A, Carninci P. 2012. Regulatory elements: Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**(4): 233-245.

Levy M, Futerman AH. 2010. Mammalian ceramide synthases. *IUBMB Life* **62**(5): 347-356.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Nature* **276**(5688): 565-570.

Li G, Holland P. 2010. The origin and evolution of ARGFX homeobox loci in mammalian radiation. *BMC Evolutionary Biology* **10**(1): 182.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. *Annals of Applied Statistics* **5**(3): 1752-1779.

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis ARJ, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T. 1997. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15**(1): 21-29.

Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. **484**(7393): 251-255.

Lise AM, Stringa E, Woodward WA, Mello MA, Tuan RS. 2000. Embryonic Limb Mesenchyme Micromass Culture as an In Vitro Model for Chondrogenesis and Cartilage Maturation. Vol 137, pp. 359-375.

Logan M, Simon HG, Tabin C. 1998. Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity. *Development* **125**(15): 2825-2835.

Logan M, Tabin CJ. 1999. Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* **283**(5408): 1736-1739.

Luscombe NM, Austin SE, Berman HM, Thornton JM. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol* **1**(1): REVIEWS001.

MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. 2011. Genome-wide transcription factor binding: beyond direct target regulation. **27**(4): 141-148.

Mann RS, Lelli KM, Joshi R. 2009. *Chapter 3 - Hox Specificity: Unique Roles for Cofactors and Collaborators*. Elsevier Inc.

Maston GA, Evans SK, Green MR. 2006. Transcriptional Regulatory Elements in the Human Genome. *Annu Rev Genom Human Genet* **7**(1): 29-59.

McCobb D, Best P, Beam K. 1990. The differentiation of excitability in embryonic chick limb motoneurons. *The Journal of Neuroscience* **10**(9): 2974-2984.

Menke DB, Guenther C, Kingsley DM. 2008. Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. *Development* **135**(15): 2543-2553.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**(7): 621-628.

Mortlock DP, Innis JW. 1997. Mutation of HOXA13 in hand-foot-genital syndrome. *Nat Genet* **15**(2): 179-180.

Muragaki Y, Mundlos S, Upton J, Olsen BR. 1996. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**(5261): 548-551.

Nakano K, Sakai N, Yamazaki Y, Watanabe H, Yamada N, Sezaki K, Susami T, Tokunaga K, Takato T, Uchinuma E. 2007. Novel mutations of the HOXD13 gene in hand and foot malformations. *International surgery* **92**(5): 287.

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. 2003. Scanning human gene deserts for long-range enhancers. *Science* **302**(5644): 413.

Ouyang Z, Zhou Q, Wong WH. 2009. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. **106**(51): 21521-21526.

Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive Divergence of Transcription Factor Binding in <italic>Drosophila</italic> Embryos with Highly Conserved Gene Expression. *PLoS Genet* **9**(9): e1003748.

Percival-Smith A, Muller M, Affolter M, Gehring WJ. 1990a. The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *The EMBO Journal* **9**(12): 3967-3974.

Percival-Smith A, Müller M, Affolter M, Gehring WJ. 1990b. The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *The EMBO Journal* **9**(12): 3967-3974.

Perez WD, Weller CR, Shou S, Stadler HS. 2010. Survival of Hoxa13 homozygous mutants reveals a novel role in digit patterning and appendicular skeletal development. *Dev Dyn* **239**(2): 446-457.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.

Ranger AM, Gerstenfeld LC, Wang J, Kon T, Bae H, Gravallese EM, Glimcher MJ, Glimcher LH. 2000. The Nuclear Factor of Activated T Cells (Nfat) Transcription Factor Nfatp (Nfatc2) Is a Repressor of Chondrogenesis. *The Journal of Experimental Medicine* **191**(1): 9-22.

Riddle RD, Johnson RL, Laufer E, Tabin C. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**(7): 1401-1416.

Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F. 2011. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet* **43**(4): 379-386.

Saadi I, Kuburas A, Engle JJ, Russo AF. 2003. Dominant Negative Dimerization of a Mutant Homeodomain Protein in Axenfeld-Rieger Syndrome. *Molecular and Cellular Biology* **23**(6): 1968-1982.

Salsi V, Vigano MA, Cocchiarella F, Mantovani R, Zappavigna V. 2008. Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. *Developmental Biology* **317**(2): 497-507.

Salsi V, Zappavigna V. 2006. Hoxd13 and Hoxa13 directly control the expression of the EphA7 Ephrin tyrosine kinase receptor in developing limbs. **281**(4): 1992-1999.

Saunders JW. 1948. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J Exp Zool* **108**(3): 363-403.

Schier AF, Gehring WJ. 1992. Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. **356**(6372): 804-807.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18**(20): 6097-6100.

Seemann P. 2006. Zur Bedeutung des Wachstumsfaktors GDF5. In *FB Biologie, Chemie, Pharmazie*, Vol Dr. rer. nat. Free University Berlin.

Shou S, Scott V, Reed C, Hitzemann R, Stadler HS. 2005. Transcriptome analysis of the murine forelimb and hindlimb autopod. *Dev Dyn* **234**(1): 74-89.

Siatecka M, Sahr KE, Andersen SG, Mezei M, Bieker JJ, Peters LL. 2010. Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Krüppel-like factor. *Proc Natl Acad Sci USA* **107**(34): 15151-15156.

Siggers T, Duyzend MH, Reddy J, Khan S, Bulyk ML. 2011. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol* **7**.

Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ et al. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**(6): 1270-1282.

Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**(9): 613-626.

Stefflova K, Thybert D, Wilson Michael D, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams David J, Talianidis I, Marioni John C et al. 2013. Cooperativity and Rapid Evolution of Cobound Transcription Factors in Closely Related Mammals. *Cell* **154**(3): 530-540.

Stender JD, Kim K, Charn TH, Komm B, Chang KCN, Kraus WL, Benner C, Glass CK, Katzenellenbogen BS. 2010. Genome-Wide Analysis of Estrogen Receptor α DNA Binding and Tethering Mechanisms Identifies Runx1 as a Novel Tethering Factor in Receptor-Mediated Transcriptional Activation. *Molecular and Cellular Biology* **30**(16): 3943-3955.

Struhl K. 1981. Deletion mapping a eukaryotic promoter. *Proceedings of the National Academy of Sciences* **78**(7): 4461-4465.

Summerbell D. 1974. A quantitative analysis of the effect of excision of the AER from the chick limb-bud. *Journal of Embryology and Experimental Morphology* **32**(3): 651-660.

Sun X, Frierson HF, Chen C, Li C, Ran Q, Otto KB, Cantarel BM, Vessella RL, Gao AC, Petros J et al. 2005. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet* **37**(4): 407-412.

Svingen T, Tonissen KF. 2006. Hox transcription factors and their elusive mammalian gene targets. *Heredity* **97**(2): 88-96.

Swanson CI, Evans NC, Barolo S. 2010. Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Developmental Cell* **18**(3): 359-370.

Szeto DP, Rodriguez-Esteban C, Ryan AK, O'Connell SM, Liu F, Kioussi C, Gleiberman AS, Izpisúa-Belmonte JC, Rosenfeld MG. 1999. Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes & Development* **13**(4): 484-494.

Tarchini B, Duboule D. 2006. Control of Hoxd Genes' Collinearity during Early Limb Development. *Developmental Cell* **10**(1): 93-103.

Tickle C. 2005. Making digit patterns in the vertebrate limb. *Nature Reviews Molecular Cell Biology* **7**(1): 45-53.

Todeschini AL, Dipietromaria A, L'Hote D, Boucham FZ, Georges AB, Pandaranayaka PJE, Krishnaswamy S, Rivals I, Bazin C, Veitia RA. 2011. Mutational probing of the forkhead domain of the transcription factor FOXL2 provides insights into the pathogenicity of naturally occurring mutations. *Human Molecular Genetics* **20**(17): 3376-3385.

Ton CCT, Hirvonen H, Miwa H, Weil MM, Monaghan P, Jordan T, van Heyningen V, Hastie ND, Meijers-Heijboer H, Drechsler M et al. 1991. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67**(6): 1059-1074.

Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH. 1997. FOG, a Multitype Zinc Finger Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic Differentiation. *Cell* **90**(1): 109-119.

Tschopp P, Tarchini B, Spitz F, Zakany J, Duboule D. 2009. Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genetics* **5**(3): e1000398.

Turpaev KT. 2006. Role of transcription factor AP-1 in integration of cell signaling systems. *Mol Biol* **40**(6): 851-866.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**(suppl 2): W71-W74.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. **10**(4): 252-263.

Veitia RA. 2007. Exploring the Molecular Etiology of Dominant-Negative Mutations. *The Plant Cell Online* **19**(12): 3843-3851.

Villavicencio-Lorini P, Kuss P, Friedrich J, Haupt J, Farooq M, Türkmen S, Duboule D, Hecht J, Mundlos S. 2010. Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation. *Journal of Clinical Investigation* **120**(6): 1994-2004.

Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* **40**(2): 158-160.

Vokes SA, Ji H, Wong WH, McMahon AP. 2008. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes & Development* **22**(19): 2651-2663.

Wang B, Xu B, Cheng Z, Zhou X, Wang J, Yang G, Cheng L, Yang J, Ma X. 2012a. A novel non-synonymous mutation in the homeodomain of HOXD13 causes synpolydactyly in a Chinese family. *Clinica Chimica Acta* **413**(13-14): 1049-1052.

Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y et al. 2012b. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research* **22**(9): 1798-1812.

Whitfield TW, Wang J, Collins PJ, Partridge EC, Aldred SF, Trinklein ND, Myers RM, Weng Z. 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol* **13**(9): R50.

Wilkie AO, Tang Z, Elanko N, Walsh S, Twigg SR, Hurst JA, Wall SA, Chrzanowska KH, Maxson RE. 2000. Functional haploinsufficiency of the human homeobox gene MSX2 causes defects in skull ossification. *Nat Genet* **24**(4): 387-390.

Woltering JM, Duboule D. 2010. The Origin of Digits: Expression Patterns versus Regulatory Mechanisms. *Dev Cell* **18**(4): 526-532.

Ye T, Krebs AR, Choukrallah M-A, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Research* **39**(6): e35.

Zakany J, Duboule D. 1996. Synpolydactyly in mice with a targeted deficiency in the HoxD complex. *Nature* **384**(6604): 69-71.

-. 2007. The role of Hox genes during vertebrate limb development. *Development* **17**(4): 359-366.

Zakany J, Kmita M, Duboule D. 2004. A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* **304**(5677): 1669-1672.

Zhang Y, Larsen CA, Stadler HS, Ames JB. 2011. Structural Basis for Sequence Specific DNA Binding and Protein Dimerization of HOXA13. *PLoS ONE* **6**(8): e23069.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

Zhao C, Dave V, Yang F. 2000. Target selectivity of bicoid is dependent on nonconsensus site recognition and protein-protein interaction. *Molecular and Cellular Biology* **20**: 8772-8123.

Zhao X, Sun M, Zhao J, Leyva JA, Zhu H, Yang W, Zeng X, Ao Y, Liu Q, Liu G et al. 2007. Mutations in HOXD13 underlie syndactyly type V and a novel brachydactyly-syndactyly syndrome. *The American Journal of Human Genetics* **80**(2): 361-371.

Zhou X, Zheng C, He B, Zhu Z, Li P, He X, Zhu S, Yang C, Lao Z, Zhu Q et al. 2013. A novel mutation outside homeodomain of HOXD13 causes synpolydactyly in a Chinese family. *Bone* **57**(1): 237-241.

7 Acknowledgements

Prof. Dr. Stefan Mundlos danke ich für die Überlassung dieser Arbeit, die fortwährend interessierte Begleitung, wissenschaftliche Diskussionen über Themen der Arbeit, Gene und Genome. Weiter gilt mein Dank den Gutachtern, Prof. Dr. Petra Seemann für ihre Betreuung und Unterstützung bei hühnerbezogenen und anderen Experimenten, sowie Prof. Dr. Uwe Ohler für seine spontane Bereitschaft, diese Arbeit zu begutachten.

Mein wirklich ganz besonderer Dank gilt Dr. Jochen Hecht; zunächst dafür, zu fragen ob ich nicht eine Doktorarbeit in seinem Team anfangen möchte (eine gute Idee!), für seine Betreuung, dafür Experimente *einfach* funktionieren zu lassen wenn man dachte, dass es nicht geht, für wissen- und unwissenschaftliches Gequatsche und noch mehr. Ich habe viel gelernt.

Auch möchte ich mich sehr bei Peter Hansen bedanken, meinem *Partner in Crime* am bioinformatischen Ende der Arbeit. Sein Engagement hat es mir als Biologe ermöglicht, die ChIP-seq Experimente auszuwerten ohne ein Script schreiben zu müssen. Nach stundenlangen Telefonaten, U-Bahn- und Autofahrten weiß ich jetzt, was die diversen Algorithmen berechnen und wie weit man manchen über den Weg trauen sollte.

Meinen drei Laborkolleginnen, die mich ertragen haben: Danke Hendrikje, mit der ich das Leben als Doktorand/in und manche Probleme teilen durfte und mit der ich wirklich immer einer Meinung war. Asita, gute Seele, Organisatorin, passionierte S-Bahn-Fahrerin und meine persönliche Proteinpäpstin. Ivana, die insgeheim perfektes Deutsch spricht und eine tolle Büronachbarin ist. Till und Mareen möchte ich für Stunden um Stunden über offenen Hühnereiern danken – eine unerwartet unterhaltsame Tätigkeit. Sie können am besten die sich ins unermessliche steigernde Begeisterung verstehen, wenn im Laufe der Woche mehr und mehr missglückte Injektionen aussortiert werden.

Liebe Kollegen am MPI: Martin, Malte, Dario, Saniye, Julia, Katerina undsoweiterundsofort – ihr seid ne super Gruppe und es macht echt Spaß mit euch zu arbeiten!

Zuguterletzt möchte ich denen danken, die mein Leben über die vergangenen vier Jahre hinaus erfüllen und verbessern ; meinen Eltern, meiner Familie und Freunden im Norden, im Osten und im Westen. Ich bin froh, dass Ihr es seid, die dabei sind! Den Naphtas Paul, David, Björn und Utz für sehr sehr gute Momente, Stunden und Konzerte. Ich würde sofort alles wieder genauso machen damit es ganz anders wird. Und natürlich und ganz besonders: Danke Susanne! Ich freu mich.

8 Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel und Literaturquellen angefertigt habe.

Ich besitze bisher keinen entsprechenden Doktorgrad und habe mich nicht anderwärts um den akademischen Grad „Doctor rerum naturalium“ beworben. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Ich erkläre die Zurkenntnisnahme der Promotionsordnung der Mathematisch-Naturwissentschaftlichen Fakultät I der Humboldt-Universität zu Berlin.

_____

Berlin, den

9 Scientific Publications

Conference Contributions

2013 June 7-9	The 6th Berlin Summer Meeting, Berlin, Germany *Poster:* Functional Characterisation of Transcription Factor Mutations using ChIP-seq
2012 September 22-25	The 4th EMBO Meeting – Nice, France *Poster:* Functional Characterisation of Transcription Factor Mutations using ChIP-seq
2012 June 23-26	European Human Genetics Conference – Nürnberg, Germany *Poster:* Global Shift in Genomic Binding Profile of a Limb Malformation Associated HOXD13 Mutation
2008 July 11-12	International Conference on Epigenetics – Berlin, Germany

Publications

Spielmann M, Brancati F, Krawitz PM, Robinson PN, Ibrahim DM, Franke M, Hecht J, Lohan S, Dathe K, Nardone AM, Ferrari P, Landi A, Wittler L, Timmermann B, Chan D, Mennen U, Klopocki E and Mundlos S. "Homeotic Arm-to-Leg Transformation Associated with Genomic Rearrangements at the Pitx1 Locus." 91, no. 4 (2012): 629-635.

Ibrahim DM, Biehs B, Kornberg TB, and Klebes A. "Microarray Comparison of Anterior and Posterior Drosophila Wing Imaginal Disc Cells Identifies Novel Wing Genes." *G3: Genes|Genomes|Genetics* 3, no. 8 (2013): 1353-1362.

Ibrahim DM, Hansen P, Rödelsperger C,Stiege AC, Doelken SC, Horn D, Jäger M, Janetzki C, Krawitz P, Leschik G, Wagner F, Scheuer T, Schmidt-von Kegler M, Seemann P, Timmermann B, Robinson PN, Mundlos S and Hecht J. "Distinct Global Shifts in Genomic Binding Profiles of Limb Malformation-Associated Hoxd13 Mutations." *Genome Research*, (2013): 23(12):2091-102

Kuss P, Kraft K, Stumm J, Ibrahim DM, Vallecillo-Garcia P, Mundlos S and Stricker S. "Regulation of Cell Polarity in the Cartilage Growth Plate and Perichondrium of Metacarpal Elements by Hoxd13 and Wnt5a." *Developmental Biology*, (2014): 385(1):83-93